Cochrane Library

Implementation of treatment guidelines for specialist mental health care (Review)

Bighelli I, Ostuzzi G, Girlanda F, Cipriani A, Becker T, Koesters M, Barbui C

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Implementation of treatment guidelines for specialist mental health care

Irene Bighelli[1], Giovanni Ostuzzi[2], Francesca Girlanda[3], Andrea Cipriani[4], Thomas Becker[3], Markus Koesters[3], Corrado Barbui[5]

[1]Department of Psychiatry and Psychotherapy, Klinikum rechts der Isar, Technische Universität München, Munich, Germany. [2]Department of Neuroscience, Biomedicine and Movement Sciences, Section of Psychiatry, University of Verona, Verona, Italy. [3]Department of Psychiatry II, Ulm University, Guenzburg, Germany. [4]Department of Psychiatry, University of Oxford, Oxford, UK. [5]Department of Neuroscience, Biomedicine and Movement Sciences, Section of Psychiatry, University of Verona, Verona, Italy

Contact address: Irene Bighelli, Department of Psychiatry and Psychotherapy, Klinikum rechts der Isar, Technische Universität München, Ismaningerstr. 22, Munich, Germany. irene.bighelli@tum.de.

A B S T R A C T

Background

A huge gap exists between the production of evidence and its uptake in clinical practice settings. To fill this gap, treatment guidelines, based on explicit assessments of the evidence base, are commonly used in several fields of psychiatry, including schizophrenia and related psychotic disorders. However, it remains unclear whether treatment guidelines have any material impact on provider performance and patient outcomes, and how implementation should be conducted to maximise benefit.

Objectives

The primary objective of this review was to examine the efficacy of guideline implementation strategies in improving process outcomes (performance of healthcare providers) and patient outcomes. We also explored which components of different guideline implementation strategies could influence them.

Search methods

We searched the Cochrane Schizophrenia Group Register (March 2012 and August 2015), as well as references of included studies.

Selection criteria

Studies that examined schizophrenia-spectrum disorders to compare guideline implementation strategies with usual care or to assess the comparative efficacy of different guideline implementation strategies.

Data collection and analysis

Review authors worked independently and in duplicate to critically appraise records from 990 studies; six individual studies met the inclusion criteria. Among the six included studies, significant heterogeneity was found in the focus of the guideline, target of the intervention, implementation strategy, and outcome measures, so meta-analysis was carried out for antipsychotic co-prescribing only.

Main results

This review now includes six studies, with a total of 1727 participants. Of the six included studies, practitioner impact was assessed in four. Overall, risk of bias was rated as low or unclear, and all evidence in the 'Summary of findings' tables was graded as low or very low quality. Meta-analysis revealed that a combination of several guideline dissemination and implementation strategies targeting healthcare

professionals did not reduce antipsychotic co-prescribing in schizophrenia outpatients (2 RCTs, N = 1082, RR 1.10 CI 0.99 to 1.23; corrected for cluster design: N = 310, RR 0.97, CI 0.75 to 1.25, very low-quality evidence). One trial, which studied a nurse-led intervention aimed at promoting cardiovascular disease screening, found a significant effect in the proportion of people receiving screening (Framingham score: N = 110, RR 0.69, 95% CI 0.55 to 0.87), although in the analysis corrected for cluster design, the effect was no longer statistically significant (N = 38, RR 0.71, 95% CI 0.48 to 1.03, very low-quality evidence).

One trial reported the patient outcomes of global state, satisfaction with care, treatment adherence, and drug attitude; no effect between treatments was seen. Quality of life was not reported by any of the studies.

One trial, which studied the use of re-written guideline text compared to original text, did not find a significant effect on staff receiving training (N = 68, RR 1.03, 95% CI 0.87 to 1.21, low-quality evidence), staff receiving supervision (N = 68, RR 0.86, 95% CI 0.64 to 1.17, low-quality evidence), or staff providing psychological interventions (N = 68, RR 0.86, 95% CI 0.62 to 1.18, low-quality evidence).

Regarding participant outcomes, only one trial assessed the efficacy of a shared decision-making implementation strategy and found no impact on psychopathology, satisfaction with care, or drug attitude. Another single trial studied a multifaceted intervention to promote medication adherence and found no effect on adherence rates.

Authors' conclusions

Considering the available evidence, it is not possible to arrive at definitive conclusions. The preliminary pattern of evidence suggests that uncertainty remains about clinically meaningful and sustainable effects of treatment guidelines on patient outcomes and how best to implement such guidelines for maximal benefit.

PLAIN LANGUAGE SUMMARY

Implementation of treatment guidelines in mental health care

Background

During the past few decades, a wide range of therapies and interventions for mental health have been developed that have been supported by evidence from research. This includes research evidence on the effectiveness of pharmacological treatments (such as antipsychotic drugs) and psychological therapies (such as cognitive behavioural therapy, family therapy and psychoeducation). However, research evidence is not easily translated into practice and the everyday working of healthcare services. A huge gap exists between the production of research evidence (what is known) and its uptake in healthcare settings (what is done). Better uptake of research evidence can be achieved by increasing awareness that such evidence exists.

One method of encouraging better uptake is the use of treatment guidelines based on assessments of research evidence. Treatment guidelines are now commonly used in healthcare settings, including those providing treatment for schizophrenia. However, it remains unclear whether treatment guidelines have any positive impact on the performance of mental health services, or whether they improve outcomes for patients (such as better quality of life, improved mental state, employment, and fewer admissions to hospital).

Searches

This review is based on a search, carried out by the information specialist of the Cochrane Schizophrenia Group, in March 2012 and updated in August 2015.

Study characteristics

This review includes six studies that examine the effectiveness of guideline implementation strategies in improving healthcare services and outcomes for people with mental illness compared with the usual care provided.

Results

Although single studies provided initial evidence that implementation of treatment guidelines may achieve small changes in mental health practice, most results showed no effect.

Conclusions

With such a small number of studies, and with all main results graded by review authors as providing very low-quality evidence, it was not possible to arrive at concrete and definite conclusions. A gap in knowledge still exists about how implementation strategies might improve patient outcomes and health services. This leaves scant information for people with mental health problems, health professionals, and policy makers. More large-scale, well-designed and well-conducted studies are necessary to fill this gap in knowledge.

This plain language summary has been written by a consumer Ben Gray: Service User and Service User Expert, Rethink Mental Illness.

SUMMARY OF FINDINGS

Summary of findings for the main comparison.   Active education + Support for implementation compared with Routine care or Passive dissemination for participants with schizophrenia and related psychosis

Active education plus support for implementation compared with routine care or passive dissemination for participants with schizophrenia and related psychosis

Patient or population: participants with schizophrenia and related psychosis
Settings: specialist mental health care
Intervention: active education + support for implementation
Comparison: routine care or passive dissemination

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Routine care or Passive dissemination	Active education + Support for implementation				
Process outcome, Practitioner Impact 1. Polypharmacy at follow-up (corrected for design effect) Follow-up: 5 to12 months	**441 per 1000**	**428 per 1000** (331 to 552)	**RR 0.97** (0.75 to 1.25)	310 (2 studies)	⊕⊝⊝⊝ **very low** [1,2]	
Process outcome, Practitioner Impact: 2. Not screened for cardiovascular risk, Framingham score (corrected for design effect) Follow-up: 6 months	**895 per 1000**	**635 per 1000** (429 to 922)	**RR 0.71** (0.48 to 1.03)	38 (1 study)	⊕⊝⊝⊝ **very low** [1,3]	
Patient outcome: Global state: average score (PANSS total, high = poor) Follow-up: 18 months		Mean global state—PANSS total score—design effect corrected in the intervention groups was **1.30 lower** (10.52 lower to 7.92 higher)		59 (1 study)	⊕⊝⊝⊝ **very low** [4,5]	MD -1.30 (95% -10.52 to 7.92)
Patient outcome: satisfaction with care: average score (ZUF8, high = better satisfaction) Follow-up: 18 months		Mean satisfaction with care—ZUF8—design effect corrected in the intervention groups was **0.10 higher** (1.96 lower to 2.16 higher)		46 (1 study)	⊕⊝⊝⊝ **very low** [4,5]	MD 0.10 (95% CI -1.96 to 2.16)

Patient outcome: treatment adherence: not adherent at follow-up (corrected for design effect) Follow-up: 6 months	**385 per 1000**	**346 per 1000** (169 to 712)	**RR 0.90** (0.44 to 1.85)	52 (1 study)	⊕⊖⊖⊖ **very low** 5,6
Patient outcome: drug attitude: average score (DAI, high = positive attitude) Follow-up: 18 months		Mean drug attitude—DAI—design effect corrected in the intervention groups was **1.40 lower** (3.38 lower to 0.58 higher)		32 (1 study)	⊕⊖⊖⊖ **very low** 4,5
Patient outcome: quality of life				0 (0)	See comment	No trial reported this outcome

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval; **RR:** Risk ratio.

GRADE Working Group grades of evidence.
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1]Risk of bias: downgraded by 2 - 'very serious': Randomisation and allocation poorly described.
[2]Imprecision: downgraded by 1 - 'serious': Only two studies with a pooled treatment estimate ranging from substantial beneficial effect to substantial harmful effect.
[3]Imprecision: downgraded by 1 - 'serious': Only one study with few cases and events.
[4]Risk of bias: downgraded by 1 - 'serious': Groups were not well balanced in terms of length of hospitalisation and psychopathology ratings.
[5]Imprecision: downgraded by 2 - 'very serious': Only one study with few cases.
[6]Risk of bias: downgraded by 1 - 'serious': Groups were not well balanced in terms of ethnic groups and psychopathology ratings.

Summary of findings 2.   Re-written guideline text compared to original guideline text for patients with schizophrenia and related psychosis

Re-written guideline text compared to original guideline text for patients with schizophrenia and related psychosis
Patient or population: patients with schizophrenia and related psychosis **Settings:** specialist mental health care **Intervention:** re-written guideline text **Comparison:** original guideline text

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Original guideline text	Re-written guideline text				
Process outcome, Practitioner Impact 3. Staff not trained (as suggested by the guideline) Follow-up: 1 month	886 per 1000	912 per 1000 (771 to 1000)	**RR 1.03** (0.87 to 1.21)	68 (1 study)	⊕⊕⊖⊖ **low** [1]	
Process outcome, Practitioner Impact 4. Staff not supervised (as suggested by the guideline) Follow-up: 1 month	771 per 1000	663 per 1000 (494 to 903)	**RR 0.86** (0.64 to 1.17)	68 (1 study)	⊕⊕⊖⊖ **low** [1]	
Process outcome, Practitioner Impact 5. Staff did not provide psychological interventions (as suggested by the guideline) Follow-up: 1 month	743 per 1000	639 per 1000 (461 to 877)	**RR 0.86** (0.62 to 1.18)	68 (1 study)	⊕⊕⊖⊖ **low** [1]	
Patient outcome: Global state				0 (0)	See comment	No trial reported this outcome
Patient outcome: Satisfaction with care				0 (0)	See comment	No trial reported this outcome
Patient outcome: Treatment adherence				0 (0)	See comment	No trial reported this outcome
Patient outcome: Drug attitude				0 (0)	See comment	No trial reported this outcome
Patient outcome: Quality of life				0 (0)	See comment	No trial reported this outcome

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval; **RR:** Risk ratio;

GRADE Working Group grades of evidence

High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

---

[1] Imprecision: downgraded by 2 - 'very serious': Only one study with few cases.

Trusted evidence.
Informed decisions.
Better health.

BACKGROUND

Description of the condition

During the past few decades, a wide range of therapeutic interventions, backed by some randomised evidence, have been developed for people with schizophrenia and related psychotic disorders. These interventions include pharmacological treatment with first- and second-generation antipsychotic agents, cognitive behavioural therapy and other psychological interventions, family interventions and psychoeducation, social skills training, vocational rehabilitation, and other psychosocial rehabilitation techniques (NICE 2014). However, under ordinary circumstances these evidence-based interventions are not easily translated into practice, as access to and use of the evidence base are not straightforward for most healthcare providers in most countries of the world (Grol 2003; Grol 2008). As a consequence, a huge gap between the production of evidence (what is known) and its uptake in practice settings (what is actually done) has been repeatedly highlighted and described in different countries and in different healthcare systems (Sederer 2009).

Description of the intervention

In recent years, to promote the use of evidence, new methodologies for aggregating, synthesising, and grading the quality of evidence extracted from systematic reviews have progressively been developed, and approaches for creating clinical practice guidelines based on explicit assessments of the evidence base are commonly employed in several fields of medicine, including schizophrenia and related psychotic disorders (Barbui 2010; Barbui 2011; WHO 2010; WHO 2015). It is interesting to note that although the pathway from evidence generation to evidence synthesis and guideline development is highly developed and quite sophisticated, the pathway from evidence-based guidelines to evidence-based practice is much less developed (Cipriani 2014; Fretheim 2006). The key issues are (1) whether guidelines may have any impact on healthcare provider performance, and thus on patient outcomes and (2) how implementation should be conducted to maximise benefit (Gagliardi 2012). This is particularly relevant for those involved in producing and delivering evidence-based recommendations, including international organisations such as the World Health Organization (WHO), scientific bodies such as the World Psychiatric Association (WPA) or the American Psychiatric Association (APA), national institutes such as the UK National Institute for Health and Care Excellence (NICE), and for those with responsibilities in delivering high-quality mental health care, including national and local managers of mental healthcare systems, scientific organisations, and even single healthcare professionals.

How the intervention might work

Implementation methods range from simple interventions, such as dissemination of educational materials, to more complex and multifaceted interventions, including tutorial and consultation sessions; treatment algorithms, reminder systems, and audit and feedback; and use of psychological theories to overcome obstacles (Grimshaw 2004).

Why it is important to do this review

Current knowledge on how implementation programmes should be developed is very scant. The only systematic reviews conducted to date—to our knowledge—did not focus on schizophrenia and related disorders, and included observational studies in addition to randomised evidence (Girlanda 2013; Girlanda 2016; Weinmann 2007). The most recent updated search of this review in August 2015 identified one new eligible study. Mental health systems that set a commitment to evidence-based practice as a policy priority need to know urgently whether guidelines may have any impact on healthcare provider performance and on patient outcomes, and how implementation plans should be developed to maximise benefit at sustainable costs (Tansella 2009).

OBJECTIVES

The primary objective of this paper was to update our previous Cochrane review and examine the efficacy of guideline implementation strategies in improving process outcomes (performance of healthcare providers) and patient outcomes (Barbui 2014). We also explored which components of different guideline implementation strategies could influence them.

METHODS

Criteria for considering studies for this review

Types of studies

We included randomised controlled trials only. When a trial did not report randomisation but was described as double-blind, and the demographic details of each group were similar, we considered the study to be randomised. We excluded quasi-randomised studies, such as those allocated by using alternate days of the week. We included studies that used cluster randomisation (such as randomisation by clinician or practice). We included studies published in all languages.

Types of participants

We included adults, however defined, with schizophrenia or related severe mental disorders, including schizophreniform disorder, schizoaffective disorder, and delusional disorder. We omitted studies in non-adult populations because of the differences in medical decision-making for children and adolescents, including the parent and guardian role. As we were interested in making sure that information was relevant to the care of individuals with severe mental disorders in specialist settings, we only included studies with participants recruited in mental healthcare settings.

Types of interventions

We included any active or passive guideline implementation strategy. We defined guidelines as systematically developed statements (or algorithms, flow charts, or tables), prepared to assist decisions about appropriate health care for specific clinical circumstances. We defined implementation as any planned process and systematic introduction of guidelines with the aim of giving them a structural place in professional practice. We included passive strategies, such as guideline distribution. We classified interventions according to a taxonomy developed by the Cochrane Effective Practice and Organisation of Care Review Group (EPOC).

We included the following comparisons.

1. Guideline implementation strategy versus usual care ('no intervention' control).

2. Guideline implementation strategy A versus guideline implementation strategy B.

Types of outcome measures

We expected that outcomes would differ between studies according to the characteristics and purposes of the guideline under scrutiny. We grouped outcomes into process outcomes (performance of healthcare providers) and patient outcomes.

Primary outcomes

We considered the following process outcome.

1. Practitioner impact - as defined by each of the studies

Secondary outcomes

We considered the following patient outcomes.

1. Global state - as defined by each of the studies

2. Satisfaction with care - as defined by each of the studies

3. Treatment adherence - as defined by each of the studies

4. Drug attitude - as defined by each of the studies

5. Quality of life - as defined by each of the studies

'Summary of findings' table

We used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to interpret findings and the GRADEpro to import data from RevMan 5 to create 'Summary of findings' tables (Guyatt 2011). These tables provide outcome-specific information concerning the overall quality of evidence from each included study in the comparison, the magnitude of effect of the interventions examined, and the sum of available data on all outcomes.

We selected the following outcomes for inclusion in the 'Summary of findings' tables:

- Process outcome: Practitioner impact - as defined by individual studies
- Patient outcome: Global state - as defined by individual studies
- Patient outcome: Satisfaction with care - as defined by individual studies
- Patient outcome: Treatment adherence - as defined by individual studies
- Patient outcome: Drug attitude - as defined by individual studies
- Patient outcome: Quality of life - as defined by individual studies

For each of these, we preferred clear and clinically meaningful binary outcomes, if these were not available, we used available continuous data.

Search methods for identification of studies

Electronic searches

Cochrane Schizophrenia Group's Trials Register

In March 2012 and on 25 August 2015, the Information Specialist (CIS) searched the Cochrane Schizophrenia Group's Study-Based Register of Trials using the following search strategy, which was developed based on a literature review and in consultation with the authors of the review:

(*guideline* OR ((*Algorithm* OR *disseminat* OR *distribut* OR *health care reform* OR *health plan* OR *health polic* OR *health priorit* OR *health reform* OR *Improving care* OR *improving treatment* OR *knowledge transfer* OR *performance measure* OR *policy making* OR *professional standard* OR *research agenda* OR *research priorit* OR *research program* OR *statement* OR *treatment guid* OR *Treatment protocol*) AND (*assess* OR *evaluat* OR *Impact* OR *implement* OR *validity*))) in Title, Abstract, or Index Terms in REFERENCES OR (*Guideline*) in Intervention of STUDY

In a study-based register, searching the major concept retrieves all the synonym keywords and relevant studies because all the studies have already been organised based on their interventions and linked to the relevant topics.

The Cochrane Schizophrenia Group's Register of Trials is compiled by systematic searches of major resources (including MEDLINE, Embase, PsycINFO, PubMed, AMED, BIOSIS, CINAHL, and registries of clinical trials) and their monthly updates, handsearches, grey literature, and conference proceedings (see Group Module). There are no language, date, document type, or publication status limitations for the inclusion of records into the register.

Searching other resources

1. Reference lists

We searched all references of articles selected for inclusion for further relevant trials.

Data collection and analysis

Selection of studies

2016 update

Review authors IB and GO independently inspected citations from the searches and identified relevant abstracts. When disagreement occurred, it was resolved by discussion, and when doubt persisted, the full article was acquired for further inspection. When difficulties or disputes arose, they were resolved by discussion with a third review author (CB). Where it was not possible to resolve disagreement by discussion, we attempted to contact the authors of the study for clarification.

Previous version

Two review authors (CB and FG) independently inspected all abstracts of studies identified as described above, and selected potentially relevant reports. To ensure reliability, another review author (EA) independently inspected all abstracts. When disagreement occurred, it was resolved by discussion, and when doubt persisted, the full article was acquired for further inspection. We acquired the full articles of relevant reports for reassessment and carefully inspected them for a final decision on inclusion. CB and FG were not blinded to the names of authors, institutions, or journals of publication. When difficulties or disputes arose, they were resolved by discussion with a third review author (MK).

Data extraction and management

1. Data extraction

2016 update

IB and GO independently extracted data from the newly included study. In case of disagreement, a final decision was taken after discussion with a third review author (CB).

Previous version

Using a form for data collection, CB and FG extracted data from all included studies. To ensure reliability, MK independently extracted data from these studies. Again, any disagreement was discussed, decisions documented, and if necessary, we contacted the authors of studies for clarification. With any remaining problems, AC helped to clarify issues, and the final decisions were documented.

2. Data management

2.1 Scale-derived data

We included continuous data from rating scales only if: (a) the psychometric properties of the measurement instrument had been described in a peer-reviewed journal (Marshall 2000); and (b) the measurement instrument was not written or modified by one of the trialists for that particular trial.

2.2 Endpoint versus change data

There are advantages of both endpoint and change data. Change data can remove a component of between-person variability from the analysis. On the other hand, calculation of change needs two assessments (baseline and endpoint), which can be difficult in unstable and difficult to measure conditions, such as schizophrenia. We decided primarily to use endpoint data, and only used change data if the former were not available. We planned to combine endpoint and change data in the analysis as we preferred to use mean differences (MD) rather than standardised mean differences (SMD) throughout (Higgins 2011).

2.3 Skewed data

Continuous data on clinical and social outcomes often are not normally distributed. To avoid the pitfall of applying parametric tests to non-parametric data, we aimed to apply the following standards to all data before inclusion: when a scale starts from the finite number zero, the standard deviation (SD), when multiplied by two, is less than the mean (as otherwise, the mean is unlikely to be an appropriate measure of the centre of distribution (Altman 1996)); if a scale starts from a positive value (such as the Positive and Negative Syndrome Scale (PANSS, which can have values from 30 to 210), we planned to modify the calculation described above to take the scale starting point into account. In these cases, skew is present if 2 SD > (S – S min), where S is the mean score and S min is the minimum score. Endpoint scores on scales often have a finite starting point and ending point, and these rules can be applied. When continuous data are presented on a scale that includes the possibility of negative values (such as change data), it is difficult to tell whether data are skewed. We planned to enter skewed data from studies of fewer than 200 participants into additional tables rather than into an analysis. Skewed data pose less of a problem when means are examined if the sample size is large, and we entered such data into the syntheses.

2.4 Common measure

To facilitate comparison between trials, we had planned to convert variables that could be reported in different metrics, such as days in hospital (mean days per year, per week, or per month), to a common metric (e.g. mean days per month).

2.5 Conversion of continuous to binary

Where possible, efforts were made to convert outcome measures to dichotomous data. This could be done by identifying cut-off points on rating scales, and dividing participants accordingly into 'clinically improved' or 'not clinically improved'. It is generally assumed that if there has been a 50% reduction in a scale-derived score, such as the Brief Psychiatric Rating Scale (BPRS; Overall 1962), or the PANSS (Kay 1987), this could be considered to be a clinically significant response (Leucht 2005a; Leucht 2005b).

2.6 Direction of graphs

Where possible, we entered data so that the area to the left of the line of no effect indicated a favourable outcome for implementation strategies.

Assessment of risk of bias in included studies

2016 update

Review authors IB and GO independently assessed the risk of bias of each trial using the Cochrane 'Risk of bias' tool (Higgins 2011). If the raters disagreed, the final rating was made by consensus, with the involvement of a third review author (CB).

Previous version

Review authors CB and FG independently assessed the risk of bias of each trial using the Cochrane 'Risk of bias' tool (Higgins 2011). This set of criteria is based on evidence of associations between overestimate of effect and high risk of bias of the article, such as sequence generation, allocation concealment, blinding, incomplete outcome data, and selective reporting. If the raters disagreed, the final rating was made by consensus, with the involvement of MK. Where inadequate details of randomisation and other characteristics of trials were provided, we contacted the authors of the studies to obtain further information. We reported non-concurrence in quality assessment, but if disputes arose about which category a trial was to be allocated, resolution was made by discussion with MK.

Measures of treatment effect

1. Binary data

For binary outcomes, we calculated a standard estimation of the risk ratio (RR) and its 95% confidence interval (CI). It has been shown that RR is more intuitive than odds ratios (Boissel 1999), and that odds ratios tend to be interpreted as RR by clinicians (Deeks 2000). For statistically significant results, we used 'Summary of findings' tables to calculate the number needed to treat for an additional beneficial outcome or an additional harmful outcome and its 95% CI.

2. Continuous data

We analysed continuous data using mean differences (MD) with 95% confidence intervals (CI), or standardised mean differences (SMD), where different measurement scales were used.

Unit of analysis issues

1. Cluster trials

Studies increasingly employ cluster-randomisation (such as randomisation by clinician or practice), but analysis and pooling of clustered data pose problems (Barbui 2011a). They are commonly analysed as if the randomisation was performed on the individuals rather than on the clusters. In this case, approximately correct analyses were performed by dividing the binary data (the number of participants and the number experiencing the event) as presented in a report by a 'design effect' (Higgins 2011). This was calculated by using the mean number of participants per cluster (m) and the intraclass correlation coefficient (ICC; Design effect = 1 + (m – 1) * ICC (Higgins 2011)). If the ICC was not reported, it was assumed to be 0.1 (Ukoumunne 1999). For continuous data only, the sample size was reduced; means and standard deviations remained unchanged.

2. Studies with multiple treatment groups

Where a study involved more than two treatment arms, we presented all relevant treatment arms in the comparisons. For binary data, we planned to simply add and combine the data in a two-by-two table. For continuous data, we planned to combine the data following the guidelines in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). Where the additional treatment arms were not relevant, these data were not reproduced.

Dealing with missing data

1. Overall loss of credibility

At some degree of loss of follow-up, data must lose credibility (Xia 2009). For any particular outcome, should more than 50% of data be unaccounted for, we did not reproduce these data or use them in analyses. However, if more than 50% of those in one arm of a study were lost, but the total loss was less than 50%, we marked such data with (*) to indicate that such a result may well be prone to bias.

When binary or continuous outcomes were not reported, we asked the study authors to supply the data.

2. Binary data

In the case where attrition for a binary outcome was between 0% and 50%, and where these data were not clearly described, we presented data on a 'once-randomised-always-analyse' basis (an intention-to-treat analysis). Those leaving the study early were considered to have the same rates of negative outcome as those who completed, with the exception of the outcome of death. We planned to undertake a sensitivity analysis to test how prone the primary outcomes were to change when 'completed' data only were compared with the intention-to-treat analysis using the above assumption.

When data on people who left early were carried forward and included in the efficacy evaluation (last observation carried forward, LOCF), they were analysed according to the primary studies; when these people were excluded from any assessment in the primary studies, they were considered to have the negative outcome.

3. Continuous

3.1 Attrition

In cases where attrition for a continuous outcome was between 0 and 50%, and only data from people who completed the study to that point were reported, we presented and used these.

3.2 Standard deviations

For continuous outcomes, if SDs were not reported, but an exact standard error (SE) and CIs were available for group means, and either the P value or the t value was available for differences in the mean, we calculated them according to the rules described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). When only the SE was reported, we calculated SDs by using the formula SD = SE * Square root (N) (Higgins 2011). The *Cochrane Handbook for Systematic Reviews of Interventions* presents detailed formulae for estimating SDs from P values, t or F values, CIs, ranges, or other statistics. If these formulae did not apply, we calculated the SDs according to a validated imputation method based on the SDs of the other included studies (Furukawa 2006).

Assessment of heterogeneity

1. Clinical and methodological heterogeneity

First, we considered all of the included studies to judge clinical and methodological heterogeneity, while paying due attention to any differences in types of implementation strategies and outcome measures. If inspection of studies revealed considerable heterogeneity of guideline implementation strategies and outcome measures, we planned to not perform formal meta-analyses. Any disagreement was discussed and final decisions documented.

2. Statistical heterogeneity

2.1 Visual inspection

We visually inspected graphs to investigate the possibility of statistical heterogeneity.

2.2 Using the I^2 statistic

We investigated heterogeneity between studies by considering the I^2 statistic alongside the Chi^2 test P value. The I^2 statistic provides an estimate of the percentage of inconsistency thought to be due to chance (Higgins 2003). The importance of the observed value of the I^2 depends on (i) magnitude and direction of effects and (ii) strength of evidence for heterogeneity (e.g. P value from the Chi^2 test, or a confidence interval for the I^2). We interpreted an I^2 estimate of 50% or higher, accompanied by a statistically significant Chi^2 statistic as evidence of substantial levels of heterogeneity (Higgins 2011). When substantial levels of heterogeneity were found in the primary outcome, we explored reasons for heterogeneity (Subgroup analysis and investigation of heterogeneity).

Assessment of reporting biases

1. Protocol versus full study

Reporting biases arise when the dissemination of research findings is influenced by the nature and direction of results (Egger 1997). These are described in section 10.1 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). We tried to locate protocols of included randomised trials. If the protocol was available, we compared the outcomes in the protocol and in the published report. If the protocol was not available, we compared

the outcomes listed in the methods section of the trial report with the reported results.

2. Funnel plot

We are aware that funnel plots may be useful in investigating reporting biases, but are of limited power to detect small-study effects. We planned to not use funnel plots for outcomes where there were 10 or fewer studies, or where all studies were of similar sizes. In other cases, where funnel plots were possible, we asked for statistical advice in their interpretation.

Data synthesis

As reported above (Assessment of heterogeneity), we only calculated summary measures of intervention effect for studies that assessed the impact of similar guideline implementation strategies and used similar outcome measures. If summary measures were calculated, we used a random-effects model for analyses throughout, as it takes into account differences between studies even if there is no statistically significant heterogeneity. The disadvantage of the random-effects model is that it gives added weight to the smaller of the studies, that is, those trials that are most vulnerable to bias. However, the reader is able to choose to inspect the data using the fixed-effect model.

Subgroup analysis and investigation of heterogeneity

1. Subgroup analysis

No subgroup analyses were planned.

2. Investigation of heterogeneity

If inconsistency was high, this was reported. First, we investigated whether data had been entered correctly. Second, if data were correct, we visually inspected the graph and removed outlying studies to see if homogeneity could be restored. Should this occur with no more than 10% of the data excluded, we planned to present the data. If not, we did not pool the data but discussed these issues.

Should unanticipated clinical or methodological heterogeneity be obvious, we simply stated hypotheses regarding these for future reviews or versions of this review. We pre-specified no study characteristics that may be associated with heterogeneity except the quality of the trial method. Should another characteristic of the studies be highlighted by the investigation of heterogeneity, these post hoc reasons were discussed and the data analysed and presented. However, should no reasons for the heterogeneity be clear, the final data were presented without a meta-analysis. If data were clearly heterogeneous, we reasoned that it may be misleading to quote an average value for the intervention effect.

Sensitivity analysis

No sensitivity analyses were planned.

R E S U L T S

Description of studies

See Characteristics of included studies; Characteristics of excluded studies; Characteristics of studies awaiting classification; and Characteristics of ongoing studies.
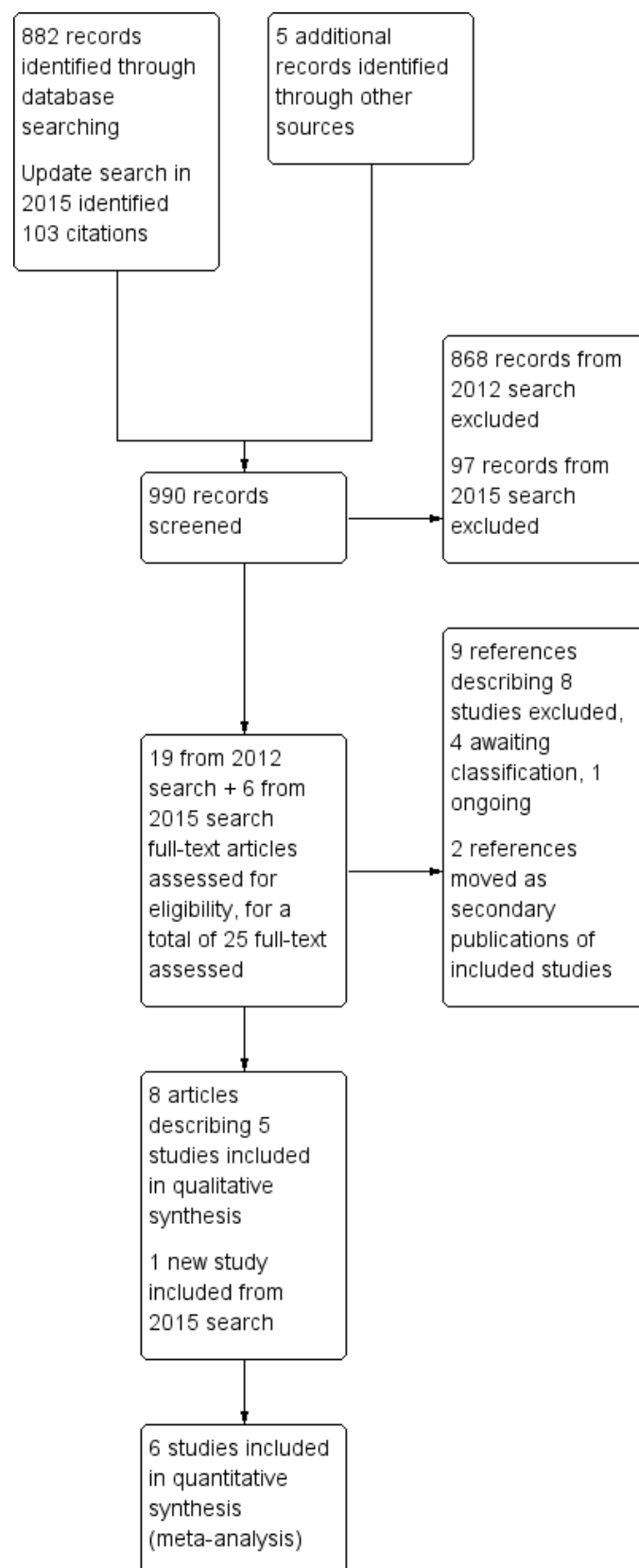
Results of the search

2016 update

The search update in 2015 yielded 103 citations, six of which were potentially relevant. One study met the inclusion criteria (Ince 2015), and another is awaiting classification (Owen 2013). See Study flow chart, Figure 1.

Figure 1. Study flow diagram.

Previous version

We inspected 882 records provided by the Cochrane Schizophrenia Group search (March 2012) and an additional five records known to us or suggested by reviewers. Of 19 potentially eligible articles, only eight, describing the results of five studies, met the inclusion criteria. See Study flow chart, Figure 1.

Included studies

2016 update

We identified one new study for inclusion in this review, for a total of six included studies.

Ince 2015 conducted a single-blind randomised trial that compared the dissemination of the original text of National Institute for Health and Clinical Excellence Guidelines guidance for schizophrenia (NICE 2014), versus dissemination of an alternative text of the same guideline, rewritten by applying good practice recommendations for producing written information. The alternative text used plain English and behaviorally specific language. Participants completed two questionnaires, at baseline and after one month. Ince 2015 is an expansion of Michie 2005, which was excluded in the previous version because it did not provide suitable outcome data. Ince 2015 added new outcome measures and therefore, was included.

Previous version

We found five studies for inclusion. Baandrup 2010 carried out a cluster-randomised comparison of a multifaceted intervention aimed at decreasing antipsychotic polypharmacy versus routine care in 602 people with schizophrenia and related psychotic disorders. They assessed prevalence of antipsychotic polypharmacy at baseline and after one year. Hamann 2006 conducted a cluster-randomised comparison of a shared decision-making intervention (printed decision aid plus planning talk) versus routine care in a sample of 107 inpatients with schizophrenia. Six wards were allocated to the experimental intervention and six to the control condition. Hudson 2008 conducted a cluster-randomised comparison of a multifaceted intervention to promote medication adherence versus basic education in six psychiatric services. They enrolled a total sample of 349 participants with schizophrenia. Osborn 2010 conducted a cluster-randomised comparison of a nurse-led intervention to improve screening for cardiovascular risk factors in people with severe mental illness. Six community mental health teams were randomly assigned to experimental (three teams) or control (three teams) conditions. A total of 121 people participated in outcome interviews. Thompson 2008 conducted a pragmatic cluster-randomised controlled trial in 19 adult psychiatric units (clusters) from the South West of England with the aim of assessing whether a multifaceted intervention was effective in reducing prescribing of antipsychotic polypharmacy.

1. Length of studies

Follow-up ranged from five to six months in three studies (Hudson 2008; Osborn 2010; Thompson 2008), to 12 to 18 months in two studies (Baandrup 2010; Hamann 2006), and lasted one month in one study (Ince 2015).

2. Participants

Participants were adults with schizophrenia (Hudson 2008; Ince 2015), schizophrenia and related psychotic disorders (Baandrup

2010; Hamann 2006; Osborn 2010), or a diagnosis of 'severe mental disorders' (Thompson 2008).

3. Setting

One study was conducted among healthcare staff from Community Mental Health Teams (Ince 2015), three studies were conducted in outpatient psychiatric settings (Baandrup 2010; Hudson 2008; Osborn 2010), and two studies recruited participants from psychiatric wards (Hamann 2006; Thompson 2008).

4. Study size

Participants numbered 349 (Hudson 2008), 121 (Osborn 2010), 107 (Hamann 2006) and 68 (Ince 2015).The other two studies carried out two cross-sectional calculations of antipsychotic polypharmacy (at baseline and at follow-up). At follow-up (primary outcome), the number of participants was 216 in the experimental group and 386 in the control group in one study (Baandrup 2010), and 220 in the control group and 260 in the experimental group in the other study (Thompson 2008).

5. Interventions and outcomes

One study, assessing the use of a rewritten version of NICE guidelines, used a self-report continuous measure to assess cognitive predictors of behaviour, and dichotomous outcomes that measured behavioural change consistent with the recommendations from the guidelines (number of participants receiving training, number of participants receiving supervision, number of participants providing psychological interventions (Ince 2015)).

Two studies assessed the efficacy of a multifaceted intervention based on existing guidelines to reduce antipsychotic polypharmacy (Baandrup 2010; Thompson 2008); prevalence of antipsychotic polypharmacy was the primary outcome in these two studies (derived from computerised medical records and from participants' medication charts). In Hamann 2006, the intervention under scrutiny consisted of written instructions to engage participants in medical decisions, and outcomes included global state, measured by the PANSS; satisfaction with care, measured with the Patient Satisfaction Questionnaire (Fragebogen zur Patientenzufriedenheit, ZUF8; Langewitz 1995); and drug attitude, measured with the Drug Attitide Inventory (DAI; Awad 1993). An enhanced implementation strategy, designed to promote guideline-concordant prescribing, was studied by Hudson 2008, who employed guideline-concordant prescribing as the primary outcome, measured by the participant's self-report of medication use over the previous 30 days and medical record abstraction. Osborn 2010 studied a nurse-led intervention aimed at promoting cardiovascular disease screening; the primary outcome of this study was the proportion of people receiving screening, as reported by participants and determined by their general practitioner (GP) notes (Osborn 2010).

Excluded studies

2016 update

We excluded four studies after careful inspection of the full text. One was not randomised; two included participants who did not have a diagnosis of schizophrenia (depression and personality disorders or dementia); the last one assessed the efficacy of a program for

the first episode of psychosis versus standard care, not guideline implementation.

The study by Michie 2005 was moved from excluded studies and included as a secondary publication of a newly included study (Ince 2015).

Previous version

We excluded five studies after careful inspection of the full text: in two, the sample included people with bipolar disorder or unipolar depression; in one, a non-randomised design was used; in another, the psychometric properties of the scale used to assess outcomes had not been validated, the participant population consisted of mental health service users with no details about diagnosis, and the outcome was a measure of cognitive determinants of implementation behaviour (Michie 2005); in the final study, the focus of the intervention was not the implementation of a guideline (see Characteristics of excluded studies).

We have created a table of suggested future reviews of schizophrenia, based on studies that we excluded from this review (Table 1).

Awaiting classification

At the 151st Annual Meeting of the American Psychiatric Association, Olfson 1998 presented the study protocol of a multi-site, prospective, controlled study, conducted to evaluate an intensive guideline implementation intervention aimed at improving the short-term outcomes of public sector participants with schizophrenia.

A publication by Owen 2013 presents the protocol of a cluster-randomised clinical trial to test effectiveness of an evidence-based quality improvement plus facilitation intervention in patients prescribed with a new antipsychotic medication. Diagnosis and setting were not clear. We contacted the authors for further information.

Study NCT00156637 was classified as ongoing in the previous version of this review. The record in clinicaltrials.gov states that the study is now completed, but the two publications indicated for this study were not consistent with the record. We contacted the authors for clarification.

Study JPRN-UMIN000004931 was classified as ongoing in the previous version of this review. The record in the UMIN Clinical Trials Registry (UMIN-CTR) states that the study is now completed, but no results were posted and the implementation strategy to promote the algorithm-based intervention was unclear. We contacted the authors for clarification.

Ongoing studies

In international repositories of trial protocols, we identified the protocol of one ongoing randomised trials that might be relevant for this review (see Characteristics of ongoing studies).

Risk of bias in included studies

We used the tool for assessment of bias described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). The overall quality of the studies was generally unclear. For an overall view of risk of bias, see Figure 2 and Figure 3.

Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.
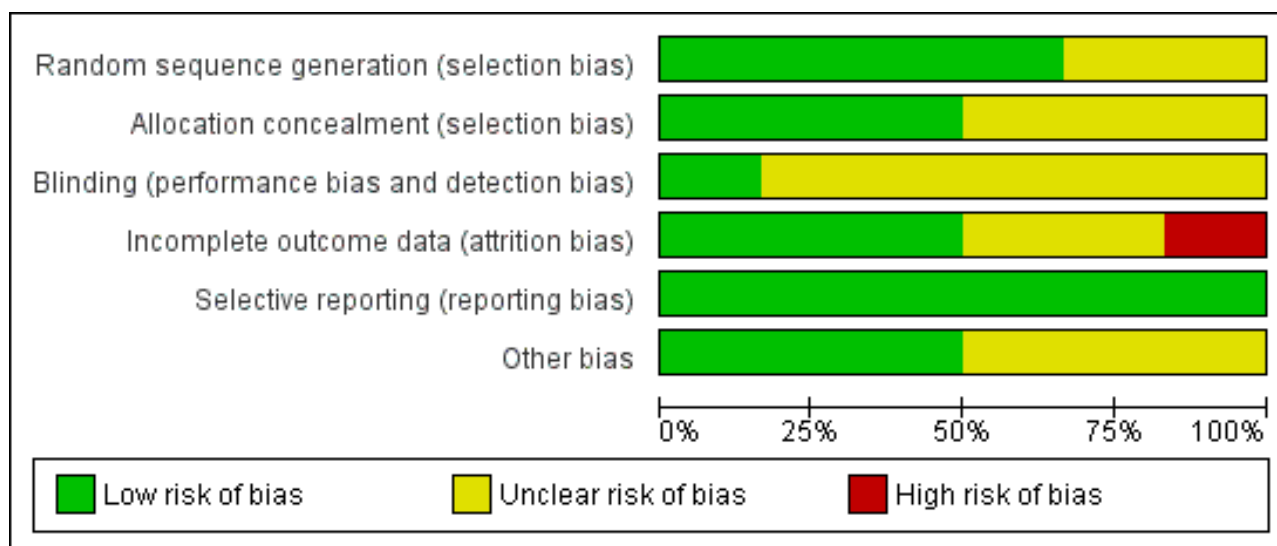
Figure 3. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

Allocation

Each included study indicated that allocation to treatment was made by random assignment. In four studies, the method used to generate a randomisation sequence was described (Hudson 2008; Ince 2015; Osborn 2010; Thompson 2008). Allocation concealment was properly described in three studies (Ince 2015; Osborn 2010; Thompson 2008).

Blinding

One study had a researcher-blind design (Ince 2015). Since the intervention was the dissemination of an alternative version of the guidelines, authors stated that it was not possible to claim that the participants were blind to the intervention, as this would depend on each individual's familiarity with the guideline. Five studies adopted an open design; this was motivated by a cluster design, which made it very difficult for study participants and researchers to remain blind to allocation (Baandrup 2010; Hamann 2006; Hudson 2008; Osborn 2010; Thompson 2008). Because of lack of information, it was very difficult to make a judgement about whether lack of blindness had an impact on the conduct or outcome of studies.

Incomplete outcome data

A high attrition rate was reported in only one study, with data available for only 66% of the initial sample at follow-up (Hamann 2006). In the other included studies, a lower attrition rate was reported.

Selective reporting

We noted that all study measures mentioned in the Methods were included in the analysis and were reported in the Results.

Other potential sources of bias

In three of the five cluster trials, as random allocation did not occur at the level of participants, participant groups were not well balanced at baseline forsociodemographic or clinical characteristics (Hamann 2006; Hudson 2008; Thompson 2008). In the other two cluster studies, participant characteristics were comparable at baseline. Although an economic conflict of interest seems not to be relevant in these studies, an intellectual bias could not be excluded, as authors of the trials were also involved in the development of the implementation strategies.

Effects of interventions

See: **Summary of findings for the main comparison** Active education + Support for implementation compared with Routine care or Passive dissemination for participants with schizophrenia and related psychosis; **Summary of findings 2** Re-written guideline text compared to original guideline text for patients with schizophrenia and related psychosis

With the exception of Baandrup 2010 and Thompson 2008, in which comparable outcome measures (antipsychotic polypharmacy) were used, critical appraisal of the included studies revealed substantial heterogeneity in focus of the guideline, target of the intervention, implementation strategy, and outcome measures.

Practitioner impact, our primary outcome, was measured by polypharmacy at follow-up and not screened for cardiovascular risk, in the studies observing the degree of coherence between the object of the recommendations contained in the guidelines and the actual behaviour of the staff. These behaviours refered to antipsychotic co-prescribing, cardiovascular disease screening, receiving training, receiving supervision, and providing psychological interventions.

Pre-defined secondary outcomes (patient outcomes) were found in the studies with the exception of quality of life, which was not reported in any of the included studies.

1. Active education + Support for implementation guidelines versus Routine care or Passive dissemination

See also: Summary of findings for the main comparison.

1.1 Process outcome, Practitioner Impact 1: Polypharmacy at follow-up

Meta-analysis of two studies revealed that a combination of several guideline dissemination and implementation strategies targeting healthcare professionals did not reduce antipsychotic co-prescribing in outpatients with schizophrenia (RR 1.10, 95% CI 0.99 to 1.23; 2 RCTs, N = 1082; corrected for cluster design: RR 0.97, 95% CI 0.75 to 1.25; N = 310; very low-quality evidence; Analysis 1.1; Baandrup 2010; Thompson 2008).

1.2 Process outcome, Practitioner Impact 2: Not screened for cardiovascular risk

Osborn 2010, in which investigators studied a nurse-led intervention aimed at promoting cardiovascular disease screening,

reported a significant effect in the proportions of people receiving screening (blood pressure: N = 96, RR 0.07, 95% CI 0.02 to 0.28; cholesterol: N = 103, RR 0.46, 95% CI 0.30 to 0.70; glucose: N = 103, RR 0.53, 95% CI 0.34 to 0.82; BMI: N = 99, RR 0.22, 95% CI 0.08 to 0.60; smoking status: N = 96, RR 0.28, 95% CI 0.12 to 0.64; Framingham score: N = 110, RR 0.69, 95% CI 0.55 to 0.87). In the analysis corrected for cluster design, the effect was statistically significant for blood pressure and cholesterol only (blood pressure, corrected for cluster design: N = 33, RR 0.10, 95% CI 0.01 to 0.74; cholesterol, corrected for cluster design: N = 35, RR 0.49, 95% CI 0.24 to 0.99; glucose, corrected for cluster design: N = 35, RR 0.58, 95% CI 0.28 to 1.21; BMI, corrected for cluster design: N = 34, RR 0.18, 95% CI 0.02 to 1.37; smoking status, corrected for cluster design: N = 32, RR 0.25, 95% CI 0.06 to 1.03; Framingham score, corrected for cluster design: N = 38, RR 0.71, 95% CI 0.48 to 1.03, very low-quality evidence; Analysis 1.2).

1.3 Patient outcome: Global state: average score (PANSS total, high = poor)

Global state was measured in one study, which found no impact on psychopathology, as measured by the PANSS (total score scale: N = 105, MD -1.30, 95% CI -8.21 to 5.61; corrected for cluster design: N = 59, MD -1.30, 95% CI -10.52 to 7.92, very low-quality evidence; Hamann 2006; Analysis 1.3).

1.4 Patient outcome: Satisfaction with care: average score (ZUF8, high = better satisfaction)

Satisfaction with care was measured in one study, which found no impact on satisfaction with care, as measured by the Patient Satisfaction Questionnaire (N = 83, MD 0.10, 95% CI -1.43 to 1.63; corrected for cluster design: N = 46, MD 0.10, 95% CI -1.96 to 2.16, very low-quality evidence; Hamann 2006; Analysis 1.4).

1.5 Patient outcome: Treatment adherence: not adherent at follow-up

Treatment adherence was measured in one study (Hudson 2008). Although researchers found a 22.5% increase from baseline in the proportion of people rated as adherent in the experimental intervention, versus a 15.1% increase from baseline in the control group, at follow-up, the two groups did not differ in adherence rates (N = 349, RR 0.87, 95% CI 0.66 to 1.15; corrected for cluster design: N = 52, RR 0.90, 95% CI 0.44 to 1.85, very low-quality evidence; Analysis 1.5).

1.6 Patient outcome: Drug attitude: average score (DAI, high = positive attitude)

Drug attitude was examined in one study, which found no impact, measured by the DAI (N = 57, MD -1.40, 95% CI -2.88 to 0.08; corrected for cluster design: N = 32, MD -1.40, 95% CI -3.38 to 0.58, very low-quality evidence; Hamann 2006; Analysis 1.6).

1.7 Patient outcome: Quality of life

This outcome was not assessed in any of the included studies.

2. Use of re-written guideline text versus use of original guideline text

See also: Summary of findings 2.

2.1 Process outcome, Practitioner Impact 3. Staff not trained (as suggested by the guideline)

One study investigated the impact of using a re-written version of a guideline on practitioner behaviour, and then measured behavioural change that was consistent with the recommendations from the guideline. The study findings showed that healthcare professionals' behaviour of receiving training did not change (N = 68, RR 1.03, 95% CI 0.87 to 1.21, low-quality evidence; Ince 2015; Analysis 2.1).

2.2 Process outcome, Practitioner Impact 4. Staff not supervised (as suggested by the guideline)

Ince 2015 did not find a change in healthcare professionals' behaviour of receiving supervision (N = 68, RR 0.86, 95% CI 0.64 to 1.17, low-quality evidence; Analysis 2.2).

2.3 Process outcome, Practitioner Impact 5. Staff did not provide psychological interventions (as suggested by the guideline)

Ince 2015 did not find a change in healthcare professionals' behaviour of providing psychological interventions, as suggested by the guidelines (N = 68, RR 0.86, 95% CI 0.62 to 1.18, low-quality evidence; Analysis 2.3).

DISCUSSION

Summary of main results

The present systematic review found very limited evidence on how implementation programmes should be developed or implemented to bridge the guideline-practice gap in specialist mental health care. We only included six randomised studies, and only completed a meta-analysis for one outcome, as critical appraisal of included studies revealed substantial heterogeneity of the focus of guidelines, target of interventions, and implementation strategies. We would have expected that studies assessing the efficacy of different implementation programmes would have reported results of practitioner impact and patient outcomes, and we had planned to extract data on several of these aspects, including global state, treatment adherence, satisfaction with care, and quality of life. This expectation is reflected by the structure of the Summary of findings for the main comparison, and Summary of findings 2, where these outcomes are listed. However, studies reported either a measure of practitioner impact or a measure of participant outcome. In addition, only one study reported more than one measure of participant outcome, and none of the included studies analysed quality of life outcomes.

In summary, these studies suggest that uncertainty remains about clinically meaningful and sustainable effects on participant outcomes.

Overall completeness and applicability of evidence

1. Completeness

The identified studies were not sufficient to address the objectives of the present review. For the primary outcome, pooling of results was possible from two studies only, and for many secondary outcomes, data were too sparse to allow us to draw reasonable conclusions. This situation might change if future implementation studies focus on key process and participant outcomes, including polypharmacy, symptom improvement, treatment adherence,

satisfaction with care, and attitude towards antipsychotic drugs. We found no data at all on quality of life (Summary of findings for the main comparison; Summary of findings 2). In short, many questions about implementation of guidelines remain unanswered.

The value of other interventions that may promote guideline use, including educational activities, social engagement, clinical support systems, incentives, and audit and feedback exercises, was not directly assessed in the included studies, although in some cases, the implementation strategy included some components of these elements.

2. Applicability

Available evidence on guideline implementation in mental health care is very sparse and involves a challenging applicability issue; it may be difficult to extrapolate study findings to a context of care different from that in which the findings were generated. Also, it cannot be assumed that an implementation strategy that proved successful for implementing a specific therapeutic behaviour may be similarly successful for implementing other therapeutic behaviours.

Another challenging aspect, only partially covered by the included studies, refers to the possibility that characteristics of some guidelines may enhance their uptake in clinical practice. For example, the content and format of guidelines may facilitate or impede their use, and addressing and taking into consideration these elements in the initial phases of guideline development may increase their chance of being implemented. It would be relevant to formally test whether careful consideration of these aspects may indeed lead to better and sustained guideline implementation in mental health care.

Quality of the evidence

In five of the included studies, a cluster-design was used. A major problem with cluster trials is that the identification and recruitment of individuals occur after random allocation of the clusters has been carried out. Therefore, it is possible that investigators enrolled participants without being blind to allocation status, and this may have introduced a potential source of bias, as knowledge of whether each cluster was an intervention or control cluster could, in theory, have affected the types of participants recruited. Hence, the potential for selection bias within clusters may be relevant (Barbui 2011a). In addition, as the unit of allocation was the cluster and not the individual, comparability at baseline for individuals may not be straightforward. We noted that in four of the five cluster trials, authors reported some baseline differences in sociodemographic and clinical characteristics.

Another compelling aspect was that the characteristics of the interventions under study did not easily allow blinding of those delivering components of the intervention and those receiving the intervention. Although this inability to blind is a distinctive feature of cluster trials, it may be possible to assess outcomes blind to allocation status, for example, by employing outcome assessors who were not involved in the conduct of the study and are masked to the allocated interventions. It is unclear whether similar approaches were employed in the included studies.

A final issue was that despite our attempt to include all randomly assigned participants in the analyses, the cluster design posed

some problems in specific circumstances. For example, in Osborn 2010, six randomised clusters included total numbers of 59 and 62 participants in the experimental and control groups. However, when the primary outcome, the proportion of participants who underwent cardiovascular screening, was measured, only those who needed screening were included; this led to the exclusion of up to seven people in the experimental arm and up to 13 in the control arm. Although it seems clinically reasonable to exclude those who had already been screened, from a methodological viewpoint, it remains unclear whether these exclusions might have had a negative impact.

Potential biases in the review process

The present systematic review has limitations. A first concern is the possibility that the search strategy may have missed some studies, because publications did not use common keywords, or were labelled with subject headings of guideline implementation initiatives that we did not capture. A second concern is that the definition of 'guideline' that we used (systematically developed statements or algorithms, flow charts, and tables to assist decisions about appropriate health care for specific clinical circumstances (Barbui 2012)) inevitably left some subjectivity in deciding whether a strategy could be considered a guideline, especially when written instructions were embedded into more complex packages of care. We believe that this should not have had a major impact on the review—but such an impact remains a possibility.

Because of substantial heterogeneity of focus of the guidelines, target of the interventions, and implementation strategies, formal meta-analyses for individual effect sizes (beyond the qualitative summary of findings) was not feasible.

Agreements and disagreements with other studies or reviews

In 2004, Grimshaw and colleagues published a landmark systematic review of the effects of guideline implementation strategies; this review included 235 studies reporting 309 comparisons of guideline dissemination and implementation strategies (Grimshaw 2004). Both randomised and non-randomised studies were considered. The review authors found that although some studies suggested it was possible to change healthcare practitioner behaviours, the overall evidence base was heterogeneous instudy designs used, populations studied, implementation strategies applied, and study quality assessed. Consequently, the review authors concluded that an imperfect evidence base supported decisions about which change strategies were likely to be efficient under different circumstances. The main findings of the present review are in line with this conclusion, and suggest that similar considerations may apply to mental healthcare settings.

In the field of mental health care, the issue of whether guidelines may have an impact on doctor or practitioner performance and patient outcome was investigated in a systematic review of randomised and non-randomised studies that enrolled participants with any psychiatric disorders. This review included only 18 studies, nine of which were randomised trials (Weinmann 2007). Although 12 studies evaluated the implementation of psychiatric guidelines in primary care settings, only five studies were carried out in mental healthcare settings. Implementation methods ranged from simple interventions, such as dissemination

of educational materials, to more complex and multifaceted interventions, including tutorial and consultation sessions, and use of treatment algorithms, reminder systems, audit and feedback, and psychological theories to overcome obstacles. Analysis of these 18 studies revealed that multifaceted interventions were more likely to have an impact on doctor performance and patient outcome, albeit effect sizes were generally modest. In total, only four studies showed a significant effect on participant outcomes (Weinmann 2007).

The work by Weinmann was updated in 2013, examining the impact of clinical guideline implementation strategies in mental healthcare settings (Girlanda 2013). This review focused on studies published after 2007 only and included 13 studies, seven of which were randomised trials. Outcomes were grouped into process outcomes (rates of polypharamcy and guideline adherence) and patient outcomes (number of weeks in any episode, time to remission, treatment adherence, and remission rates). Implementation strategies included educational material, reminders, meetings, and audit and feedback. Most of the included studies suggested a trend toward an improvement of healthcare provider performance and clinical condition of patients following guideline implementation, but effects were modest. Authors concluded that there was a lack of high-quality evidence; only initial data on how psychiatric guidelines should be effectively implemented was consistent with the main findings of the present review. A subsequent review with meta-analysis included 18 studies and found limited evidence on the efficacy of guideline implementation in mental health care (Girlanda 2016).

Audit of clinical activities and feedback to doctors may be a relevant component of any implementation strategy. Knaup and colleagues, who systematically reviewed controlled studies that evaluated audit and feedback strategies, showed a positive effect on mental health outcomes, at least in the short term (Knaup 2009). This finding seems intuitive, as guideline implementation is meant to be iterative;after implementation, guideline use and outcomes should be monitored, and the findings used to inform ongoing quality improvement efforts, as the ultimate goal of any implementation activity is continuous quality improvement. We note that none of the studies included in the present systematic review used formal audit and feedback activities as part of the implementation strategies applied.

AUTHORS' CONCLUSIONS

Implications for practice

1. For people with schizophrenia

The findings of this review have few implications for people with schizophrenia. Patients and their families should continue to question clinicians about the basis of their care and the reasoning behind using a specific drug, care package, or psychological intervention. In this way, people can encourage clinicians to think about the reasons for using one particular approach rather than another, and this, in turn, could encourage use of guidelines.

2. For clinicians

Few studies assessing the impact of guideline implementation strategies for people with schizophrenia or related psychotic disorders suggested that significant changes in clinically meaningful and sustainable effects influenced participant

outcomes. It is surprising that although the pathway from evidence generation to evidence synthesis and guideline development is highly developed, the pathway from evidence-based guidelines to evidence-based practice is much less developed and is examined in only a few studies. This is very relevant for healthcare professionals, who are left with limited instructions on how to make best use of available guidelines. If such instructions are given to practitioners, they should be rolled out within the context of a real-world randomised trial of substantial size for evaluation of their worth.

3. For policy makers and funders

The present systematic review found scant and imperfect evidence to support decisions about which change strategies are likely to be efficient in mental healthcare settings. However, the following practical consideration may be implicitly derived from the existing literature (Barbui 2012a).

Treatment guidelines should be developed as locally as possible, or should be adapted locally, to take into account issues such as value judgements, resource use, local context characteristics, and feasibility, which are aspects that may be widely different in different contexts. This may have a profound impact on the likelihood that the guideline is implemented, as healthcare professionals may be reluctant to adhere to standards of care set by others. Recommendations should reflect a balanced approach between care of individual patients and how work is organised. This may be particularly relevant in mental health care because new and better interventions (e.g. early interventions for psychotic patients, assertive community treatment, community mental health interventions, vocational and rehabilitative interventions) cannot be delivered in the absence of functioning mental healthcare systems.

Existing evidence suggests that audit and feedback systems are relevant for fidelity reasons, that is, to check the degree of coherence between what is recommended and what is actually done. Audit and feedback of patient outcomes is also essential for internal accountability reasons, that is, to provide continuous feedback to professionals, who need to know the true impact of their practice, and to mental healthcare planners, who may wish to include in their decision-making process, among other considerations, local outcome data. Audit and feedback of patient outcomes may be relevant for external accountability reasons as well, that is, to provide patients, families, and the public with data that may be used in making more informed choices, and to provide feedback to science by producing processes and outcome data that may generate new research hypotheses, which may be formally tested using experimental designs.

Implications for research

1. General

We recommend that any guideline implementation programme should be described and documented, thereby increasing our knowledge on how to make best use of available evidence to improve practice. Even experiences of guidelines produced and implemented at a local level should be reported, and could provide useful information (Ostuzzi 2013). Successful and unsuccessful experiences should be given visibility, even if they have not been studied in experimental conditions, as happens in clinical medicine, where reports on single cases are given visibility to describe new clinical scenarios and new solutions.

2. Trials

If feasible, the impact of guideline implementation programmes should be studied using reliable study designs, such as randomised trials (Cipriani 2009), and cluster-randomised trials (Barbui 2011a), and a pragmatic approach to decrease the huge imbalance between what we know and what we actually do. We realise that the design of such studies takes greater care and attention to detail than is possible to provide in a review, but we have considered available data at some length, and suggest a broad outline of a design in Table 2.

3. Reviews

Most excluded studies enrolled heterogeneous participant populations, including people with affective and non-affective psychosis and unipolar depression. These studies may find their way into reviews related to the impact of guideline implementation strategies in more general populations of psychiatric patients.

A C K N O W L E D G E M E N T S

R E F E R E N C E S

References to studies included in this review

Baandrup 2010 *{published data only}*

Baandrup L, Allerup P, Lublin H, Nordentoft M, Peacock L, Glenthoj B. Evaluation of a multifaceted intervention to limit excessive antipsychotic co-prescribing in schizophrenia out-patients. *Acta Psychiatrica Scandinavica* 2010;**122**:367–74.

NCT00541398. Antipsychotic polypharmacy: prevalence, background and consequences. clinicaltrials.gov/ct2/show/ NCT00541398 (first received 9 October 2007).

Hamann 2006 *{published data only}*

Hamann J, Cohen R, Leucht S, Busch R, Kissling W. Shared decision making and long-term outcome in schizophrenia treatment. *Journal of Clinical Psychiatry* 2007;**68**:992-7.

* Hamann J, Langer B, Winkler V, Busch R, Cohen R, Leucht S, et al. Shared decision making for in-patients with schizophrenia. *Acta Psychiatrica Scandinavica* 2006;**114**:265-73.

Hudson 2008 *{published data only}*

* Hudson TJ, Owen RR, Thrush CR, Armitage TL, Thapa P. Guideline implementation and patient-tailoring strategies to improve medication adherence for schizophrenia. *Journal of Clinical Psychiatry* 2008;**69**:74-80.

Owen RR, Hudson T, Thrush T, Thapa P, Armitage T, Landes RD. The effectiveness of guideline implementation strategies on improving antipsychotic medication management for schizophrenia. *Medical Care* 2008;**46**:686-91.

Williams DK, Thrush CR, Armitage TL, Owen RR, Hudson TJ, Thapa P. The effect of guideline implementation strategies on akathisia outcomes in schizophrenia. *Journal of Applied Research* 2003;**4**:470-82.

Ince 2015 *{published data only}*

* Ince P, Tai S, Haddock G. Using plain English and behaviourally specific language to increase the implementation of clinical guidelines for psychological treatments in schizophrenia.. *Journal of Mental Health* 2015;**24**(3):129-33.

Michie S, Lester K. Words matter: increasing the implementation of clinical guidelines. *Quality and Safety Health Care* 2005;**14**:367–70.

Osborn 2010 *{published data only}*

Osborn DPJ, Nazareth I, Wright CA, King MB. Impact of a nurse-led intervention to improve screening for cardiovascular risk factors in people with severe mental illnesses. Phase-two cluster randomised feasibility trial of community mental health teams. *BMC Health Services Research* 2010;**10**(61):1-13.

Thompson 2008 *{published data only}*

Thompson A, Sullivan SA, Barley M, Strange SO, Moore L, Rogers P, et al. The DEBIT trial: an intervention to reduce antipsychotic polypharmacy prescribing in adult psychiatry wards - a cluster randomized controlled trial. *Psychological Medicine* 2008;**38**:705–15.

References to studies excluded from this review

Azrin 1998 *{published data only}*

Azrin N, Teichner G. Evaluation of an instructional program for improving medication compliance for chronically mentally ill outpatients. *Behaviour Research and Therapy* 1998;**36**:849-61.

Brunette 2015 *{published data only}*

Brunette MF. Facilitators and barriers to implementation of coordinated specialty care in U.S. community mental health clinic. *Schizophrenia Bulletin* 2015;**41**:S304.

Byng 2004 *{published data only}*

Byng R, Jones R, Leese M, Hamilton B, McCrone P, Craig T. Exploratory cluster randomised controlled trial of shared care development for long-term mental illness. *British Journal of General Practice* 2004;**54**:259-66.

De Beurs 2013 *{published data only}*

De Beurs DP, De Groot MH, Bosmans JE, De Keijser J, Mokkenstorm J, Verwey B, et al. Reducing patients' suicide ideation through training mental health teams in the application of the Dutch multidisciplinary practice guideline on assessment and treatment of suicidal behavior: study protocol of a randomized controlled trial. *Trials* 2013;**14**:372.

Koivunen 2010 *{published data only}*

Koivunen M, Valimaki M, Patel A, Knapp M, Hatonen H, Kuosmanen L, et al. Effects of the implementation of the web-based patient support system on staff's attitudes towards computers and IT use: a randomised controlled trial. *Scandinavian Journal of Caring Sciences* 2010;**24**:592–9.

Robinson 2015 *{published data only}*

Robinson DG, Schooler NR, John M, Correll CU, Marcy P, Addington J, et al. Prescription practices in the treatment of first-episode schizophrenia spectrum disorders: data from the national RAISE-ETP study. *American Journal of Psychiatry* 2015;**172**(3):237-48.

Rush 2003 *{published data only}*

Miller A, Crismon ML, Rush AJ, Chiles J, Kashner TM, Toprac MG, et al. The Texas medication algorithm project: clinical results for schizophrenia. *Schizophrenia Bulletin* 2004;**30**:627-47.

* Rush AJ, Crismon ML, Kashner TM, Toprac MG, Carmody TJ, Trivedi MH, et al. Texas medication algorithm project, Phase 3 (TMAP-3): rationale and study design. *Journal of Clinical Psychiatry* 2003;**64**:357-69.

Tjia 2014 *{published data only}*

Tjia J, Lemay C, Mazor KM, Briesacher BA, Field TS, Kanaan AO, et al. Dissemination of evidence-based antipsychotic prescribing guidelines to nursing homes: a cluster randomized trial. *Journal of the American Geriatrics Society* 2014;**62**:1289-98.

References to studies awaiting assessment

JPRN-UMIN000004931 *{published data only}*

UMIN000004931. An intervention study of optimizing algorithm-based pharmacological treatment for schizophrenia. upload.umin.ac.jp/cgi-open-bin/ctr_e/index.cgi (date of disclosure of the study information 22 January 2011).

NCT00156637 *{published data only}*

* NCT00156637. A study of strategies to improve schizophrenia treatment. clinicaltrials.gov/ct2/show/NCT00156637 (first received 7 September 2005).

Rosen CS, Tiet QQ, Harris AH, Julian TF, McKay JR, Moore WM, et al. Telephone monitoring and support after discharge from residential PTSD treatment: a randomized controlled trial. *Psychiatric Services* 2013;**64**(1):13-20.

Young AS, Niv N, Chinman M, Dixon L, Eisen SV, Fischer EP, et al. Routine outcomes monitoring to support improving care for schizophrenia: report from the VA Mental Health QUERI. *Community Mental Health Journal* 2011;**47**(2):123-35.

Olfson 1998 *{published data only}*

Olfson M. Implementing practice guidelines for schizophrenia. 151st Annual Meeting of the American Psychiatric Association; 1998 May 30-June 4; Toronto (ON). 1998.

Owen 2013 *{published data only}*

Owen RR, Drummond KL, Viverito KM, Marchant K, Pope SK, Smith JL, et al. Monitoring and managing metabolic effects of antipsychotics: a cluster randomized trial of an intervention combining evidence-based quality improvement and external facilitation. *Implementation Science* 2013;**8**:120.

References to ongoing studies

NCT00182494 *{published data only}*

NCT00182494. A five year, prospective, randomized, blinded, controlled trial comparing the efficacy of a modified diabetes prevention protocol and the standard comprehensive outpatient care in lowering the incidence of new onset diabetes among people treated for schizophrenia and are at risk to develop type II diabetes mellitus. clinicaltrials.gov/ct2/show/NCT00182494 (first received 14 September 2005).

Additional references

Altman 1996

Altman DG, Bland JM. Detecting skewness from summary information. *BMJ* 1996;**313**:1200. [MEDLINE: 1997074332]

Awad 1993

Awad AG. Subjective response to neuroleptics in schizophrenia. *Schizophrenia Bulletin* 1993;**19**:609-18.

Barbui 2010

Barbui C, Dua T, Van Ommeren M, Yasamy MT, Fleischmann A, Clark N, et al. Challenges in developing evidence-based recommendations using the GRADE approach: the case of mental, neurological, and substance use disorders. *PLOS Medicine* 2010;**7**(8):e1000322.

Barbui 2011

Barbui C, Cipriani A. What are evidence-based treatment recommendations?. *Epidemiology and Psychiatric Sciences* 2011;**20**:29-31.

Barbui 2011a

Barbui C, Cipriani A. Cluster randomised trials. *Epidemiology and Psychiatric Sciences* 2011;**20**:307-9.

Barbui 2012

Barbui C, Koesters M, Cipriani A. Implementation of treatment guidelines for specialist mental health care. *Cochrane Database of Systematic Reviews* 2012, Issue 4. [DOI: 10.1002/14651858.CD009780]

Barbui 2012a

Barbui C, Tansella M. Guideline implementation in mental health: current status and future goals. *Epidemiology and Psychiatric Sciences* 2012;**21**:227-9.

Boissel 1999

Boissel JP, Cucherat M, Li W, Chatellier G, Gueyffier F, Buyse M, et al. The problem of therapeutic efficacy indices. 3. Comparison of the indices and their use. *Therapie* 1999;**54**(4):405-11.

Cipriani 2009

Cipriani A, Geddes JR. What is a randomised controlled trial?. *Epidemiology and Psychiatric Sciences* 2009;**18**:191-4.

Cipriani 2014

Cipriani A, Furukawa TA. Advancing evidence-based practice to improve patient care. *Evidence Based Mental Health* 2014;**17**(1):1-2.

Deeks 2000

Deeks J. Issues in the selection for meta-analyses of binary data. 8th International Cochrane Colloquium; 2000 Oct 25-28; Cape Town (SA). Cochrane Collaboration, 2000.

Egger 1997

Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**:629–34.

Fretheim 2006

Fretheim A, Schünemann HJ, Oxman AD. Improving the use of research evidence in guideline development. 15. Disseminating and implementing guidelines. *Health Research Policy and Systems* 2006;**4**:27.

Furukawa 2006

Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. *Journal of Clinical Epidemiology* 2006;**59**:7–10.

Gagliardi 2012

Gagliardi AR. Translating knowledge to practice: optimizing the use of guidelines. *Epidemiology and Psychiatric Sciences* 2012;**21**(3):231-6. [DOI: 10.1017/S2045796012000212]

Girlanda 2013

Girlanda F, Fiedler I, Ay E, Barbui C, Koesters M. Guideline implementation strategies for specialist mental healthcare. *Current Opinion in Psychiatry* 2013;**26**(4):369-75.

Girlanda 2016

Girlanda F, Fiedler I, Becker T, Barbui C, Koesters M. The evidence-practice gap in specialist mental healthcare: a systematic review and meta-analysis of guideline implementation studies. British Journal of Psychiatry 2016 Jul 21 [Epub ahead of print]. [DOI: 10.1192/bjp.bp.115.179093]

GRADEpro [Computer program]

GRADE Working Group, McMaster University. GRADEpro GDT. Hamilton (ON): GRADE Working Group, McMaster University, 2014.

Grimshaw 2004

Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technology Assessment* 2004;**8**(6):1-72.

Grol 2003

Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003;**362**:1225–30.

Grol 2008

Grol 2008. Knowledge transfer in mental health care: how do we bring evidence into day-to-day practice?. *Canadian Journal of Psychiatry* 2008;**53**:275–6.

Guyatt 2011

Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction - GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology* 2011;**64**:383-94.

Higgins 2003

Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557-60.

Higgins 2011

Higgins JPT, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (update March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Kay 1987

Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin* 1987;**13**:261-76.

Knaup 2009

Knaup C, Koesters M, Schoefer D, Becker T, Puschner B. Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *British Journal of Psychiatry* 2009;**195**:15-22.

Langewitz 1995

Langewitz W, Keller A, Denz M, Wossmer-Buntschu B, Kiss A. The patient satisfaction questionnaire: a suitable tool for quality control in the physician-patient relationship?. *Psychotherapy and Psychosomatic Medicine and Psychology* 1995;**45**:351-7.

Leucht 2005a

Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel R. Clinical implications of brief psychiatric rating scale scores. *British Journal of Psychiatry* 2005;**187**:366-71. [PUBMED: 16199797]

Leucht 2005b

Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR. What does the PANSS mean?. *Schizophrenia Research* 2005;**79**(2-3):231-8. [PUBMED: 15982856]

Marshall 2000

Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *British Journal of Psychiatry* 2000;**176**:249-52.

Michie 2005

Michie S, Lester K. Words matter: increasing the implementation of clinical guidelines. *Quality and Safety Health Care* 2005;**14**:367–70.

NICE 2014

National Institute for Health and Clinical Excellence. Psychosis and Schizophrenia in Adults. Treatment and Management. The British Psychological Society and The Royal College of Psychiatrists, 2014.

Ostuzzi 2013

Ostuzzi G, Bighelli I, Carrara BV, Dusi N, Imperadore G, Lintas C, et al. Making the use of psychotropic drugs more rational through the development of GRADE recommendations in specialist mental healthcare. *International Journal of Mental Health Systems* 2013;**7**:14.

Overall 1962

Overall JE, Gorham DR. The brief psychiatric rating scale. *Psychological Reports* 1962;**10**:799-812.

RevMan 5 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Sederer 2009

Sederer LI. Science to practice: making what we know what we actually do. *Schizophrenia Bulletin* 2009;**35**:714-8.

Tansella 2009

Tansella M, Thornicroft G. Implementation science: understanding the translation of evidence into practice. *British Journal of Psychiatry* 2009;**195**:283-5.

Ukoumunne 1999

Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ. Methods for evaluating area-wide and organisation-based intervention in health and health care: a systematic review. *Health Technology Assessment* 1999;**3**(5):iii-92.

Välimäki 2012

Välimäki M, Hätönen H, Lahti M, Kuosmanen L, Adams CE. Information and communication technology in patient education and support for people with schizophrenia. *Cochrane Database of Systematic Reviews* 2012, Issue 10. [DOI: 10.1002/14651858.CD007198.pub2]

Weinmann 2007

Weinmann S, Koesters M, Becker T. Effects of implementation of psychiatric guidelines on provider performance and patient outcome: systematic review. *Acta Psychiatrica Scandinavica* 2007;**115**:420–33.

WHO 2010

World Health Organization. MhGAP Intervention Guide for Mental, Neurological and Substance Use Disorders in Non-Specialized Health Settings. Geneva: World Health Organization, 2010.

WHO 2015

World Health Organization. WHO mhGAP Guideline Update. Update of the Mental Health Gap Action Programme (mhGAP) Guideline for Mental, Neurological and Substance use Disorder. Geneva: World Health Organization, 2015.

Xia 2009

Xia J, Adams CE, Bhagat N, Bhagat V, Bhoopathi P, El-Sayeh H, et al. Losing participants before trial ends erodes credibility of finding. *Psychiatric Bulletin* 2009;**33**(7):254-7.

References to other published versions of this review

Barbui 2014

Barbui C, Girlanda F, Ay E, Cipriani A, Becker T, Koesters M. Implementation of treatment guidelines for specialist mental health care. *Cochrane Database of Systematic Reviews* 2014, Issue 1. [DOI: 10.1002/14651858.CD009780.pub2]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Baandrup 2010

Methods	Allocation: cluster-randomised trial.
	Blindness: not blind, outcome data from computerised prescription drug records.
	Duration: 12 months.
Participants	Diagnosis: schizophrenia; schizotypal, delusional and schizoaffective disorders.
	N of clusters = two municipalities, balanced for baseline, prevalence of antipsychotic polypharmacy, socioeconomic status (of population in catchment areas), and functional level of inpatients and outpatients.
	N of participants = 216 (experimental group); 386 (control group).
	Sex: 56.9% M (experimental group); 57.5% M (control group).
	Mean age, years: 38.8 (experimental group); 37.5 (control group)
	Setting: outpatients in Denmark.
Interventions	1. Multifaceted educational intervention consisting of one day (six hours) of traditional education (didactic lectures), where current evidence for the psychopharmacological treatment of schizophrenia and existing clinical guidelines were presented, with special emphasis on avoiding or reducing irrational antipsychotic polypharmacy. In addition, six times during the following year, a 3-hour session with case-based education was offered. Finally, a warning implemented in the electronic drug prescribing system popped up every time antipsychotic polypharmacy was about to be prescribed. Target of the intervention: doctors, nurses, psychologists, other professionals.

Cochrane
Library

Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

Baandrup 2010 *(Continued)*

	2. Routine care.
Outcomes	Practitioner impact: antipsychotic polypharmacy (review of computerised prescription drug records).
Notes	Mean number of participants per cluster (m): 301.
	Intraclass coefficent (ICC): not reported.
	Design effect: 31.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Two Danish municipalities were randomly assigned to intervention (Viborg) and control settings (Esbjerg).
Allocation concealment (selection bias)	Unclear risk	No details reported.
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Participants might have been aware of the group (control or intervention) to which they were allocated because of the nature of the intervention. The two clusters were situated sufficiently far away from each other that contamination was avoided.
Incomplete outcome data (attrition bias) All outcomes	Low risk	The two randomly assigned units (clusters) were included in the analysis. The study consisted of two cross-sectional surveys—one carried out at baseline and the second after one year of follow-up. The primary outcome was prevalence of polypharmacy at follow-up (cross-sectional nature of the primary outcome measure).
Selective reporting (reporting bias)	Low risk	All outcome measures reported in the Methods were described in the analysis and reported in the Results.
Other bias	Low risk	The control group and the intervention group were perfectly matched regarding prevalence of antipsychotic polypharmacy at baseline. The study was supported by the National Board of Health in Denmark (0-204-03-9/9) and the Wørzner Foundation.

Hamann 2006

Methods	Allocation: cluster-randomised trial.
	Blindness: not blind.
	Duration: 18 months.
Participants	Diagnosis: schizophrenia or schizophreniform disorder.
	N of clusters: 12 acute psychiatric wards.
	N of participants: 107.
	Sex: 59% M (experimental group); 47% M (control group).
	Mean age, years: 35.5 (experimental group); 39.6 (control group).
	Exclusion criteria: severe mental retardation, lack of fluency in German and refusal to give written informed consent.

Hamann 2006 *(Continued)*

	Setting: two German state hospitals.
Interventions	1. Printed decision aid plus planning talk. A printed decision aid was initially developed in co-operation with a number of psychiatrists, psychiatric nurses and former patients. The final version of the decision aid was a 16-page booklet covering the pros and cons of oral versus depot formulations, first- versus second-generation antipsychotics, psychoeducation and types of sociotherapeutic interventions. These booklets were presented to participants through the head nurse of the ward as soon as the psychiatrist in charge believed that participants were able to co-operate. The nurses had been trained in assisting participants to work through the booklet and in answering any requests for information. Within the decision aid, participants were asked to write down their experiences with previous antipsychotic medications and to indicate their preferences on each topic regarding the available options. Participants met with their physicians within 24 hours after working through the decision aid with their nurse so they could reach an agreement with the psychiatrist on further treatment according to the preferences indicated by the participant in the booklet. Target of the intervention: nurses and doctors.
	2. Routine care.
Outcomes	Global state: PANSS.
	Satisfaction with treatment: ZUF8.
	Drug attitude: DAI.
Notes	Mean number of participants per cluster (m): 8.9.
	Intraclass coefficient (ICC): not reported.
	Design effect: 1.79.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Quote: "Randomization was done at the level of the wards to avoid intervention and control conditions being confounded".
		Quote: "Selection of the wards was made so as to ensure that there were six pairs of wards, with one member of each pair being randomly assigned to the control or to the interventional condition".
Allocation concealment (selection bias)	Unclear risk	No details reported.
Blinding (performance bias and detection bias) All outcomes	Unclear risk	No blinding.
Incomplete outcome data (attrition bias) All outcomes	High risk	At follow-up (18 months), data were available for 66% of the initial sample.
Selective reporting (reporting bias)	Low risk	All outcome measures reported in the Methods were included in the analysis and reported in the Results.
Other bias	Unclear risk	At baseline, the two groups were not perfectly balanced in terms of length of hospitalisation and psychopathology ratings. The trial was funded by the German Ministry of Health and Social Security (217-43794-5/9) within the funding project Der Patient als Partner im medizinischen Entscheidungsprozess.

Hudson 2008

Methods	Allocation: cluster-randomised trial.
	Blindness: not blind.
	Duration: six months.
Participants	Diagnosis: schizophrenia.
	N of clusters: six Veterans Affairs Medical Centers (VAMCs).
	N of participants: 349.
	Sex: 93% M (experimental group); 95% M (control group).
	Mean age, years: 45.4 (experimental group); 46.8 (control group).
	Setting: USA.
Interventions	1. Intervention to promote medication adherence. Research nurses were instructed to follow a written protocol for the intervention and conducted a clinical interview lasting 20 to 60 minutes. During this interview, the nurse completed a checklist of nine domains of barriers to adherence, derived from an extensive literature review. Based on the barriers identified, the nurse worked with participants to select and tailor strategies that could be used to overcome that particular barrier. The nurse conducted barrier assessments at entry into the study and at each subsequent visit; for individuals who did not have mental health appointments more frequent than every six weeks, the nurse attempted to contact the study participant to conduct a barrier assessment a minimum of every six weeks throughout the six-month study period. The nurse was given a detailed manual describing the intervention protocol, which included flexible scripts and suggestions to use in conducting clinical interviews with participants and assessment of medication adherence barriers. The protocol also specified how nurses were to maintain contact with participants during the study period and how nurses could provide feedback to each physician about the participant's treatment preferences, reported adherence and adherence barriers. Target of the intervention: nurses and participants.
	2. Routine education about schizophrenia guidelines.
Outcomes	Treatment (medication) adherence: participant's self-report of medication use over the previous 30 days and medical record abstraction for the participant's visit closest to baseline and follow-up dates.
Notes	Mean number of participants per cluster (m): 58.2.
	Intraclass coefficient (ICC): not reported.
	Design effect: 6.72.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomisation was stratified by Veteran Integrated Service Network (VISN). In each VISN, two Veteran Affairs Medical Centers (VAMCs) were randomly assigned to experimental and control conditions (two clusters for each VISN). Quote: "One of the two VAMCs per VISN was randomly selected, using a random numbers table, to receive the enhanced implementation strategy, and the other site within each pair received a basic education strategy for schizophrenia guidelines".
Allocation concealment (selection bias)	Unclear risk	No details reported.

Hudson 2008 *(Continued)*

Blinding (performance bias and detection bias) All outcomes	Unclear risk	No blinding. A combination "self-report and provider assessment" strategy was employed to measure the primary outcomes (detection bias cannot be excluded, i.e. non-adherence might have been underestimated).
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Of 349 participants enrolled at the six VAMCs, 84% (N = 293) completed both baseline and six-month follow-up assessments. Quote: "Study participants lost to follow-up were more likely to be white (45% vs. 28%) and unmarried/separated (95% vs. 84%), but there were no other significant demographic differences between those who did or did not complete follow-up interviews".
Selective reporting (reporting bias)	Low risk	All outcome measures reported in the Methods were included in the analysis and reported in the Results.
Other bias	Unclear risk	At baseline, the two groups were not perfectly balanced in terms of ethnic group and psychopathology rating. Funding was provided by a grant from the VA, J1SR& D (#CPG-97-027) and the VA South Central Mental Illness Research Education and Clinical Centre.

Ince 2015

Methods	Allocation: randomised trial Blindness: single blind. Duration: one month.
Participants	Diagnosis: schizophrenia. The target of the evaluation is represented by the clinical staff from Community Mental Health Teams (CHMT) N of centres: 13 Community Mental Health Teams (CHMT) N of participants: 68. Sex: 26% M (experimental group); 33% M (control group). Mean age, years: (experimental group); (control group). Setting: UK (two Mental Health Trusts in Greater Manchester).
Interventions	1. Alternative text of National Institute for Health and Clinical Excellence Guidelines guidance for schizophrenia (NICE 2014). Quote:" The original text was rewritten applying good practice recommendations for producing written information. The text was amended to personalise the message, avoid highly technical language, use active rather than passive verbs, avoid negative or alarmist approaches, and included appropriate numerical information. Behavioural specificity was increased by the use of behaviourally specific language and production of a checklist and decision tree of specific points from the guidance." 2. Guidance for schizophrenia original text
Outcomes	Continuous outcome: Cognitive predictors of behaviour, measured by an adapted version of a questionnaire designed for previous research Dichotomous outcomes (behavioural change consistent with the recommendations from the guidelines): number of participants receiving training, number of participants receiving supervision, number of participants providing psychological interventions.
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "An independent measure, random group allocation was used. [...] Participants from the same CHMT were randomly assigned to one of the variable conditions to reduce intervention contamination." "All members of the same participating CHMT where then randomly assigned, by the use of an allocation list constructed by a random number generator and presented with a copy of the intervention in a sealed envelope".
Allocation concealment (selection bias)	Low risk	Quote: "All members of the same participating CHMT where then randomly assigned, by the use of an allocation list constructed by a random number generator and presented with a copy of the intervention in a sealed envelope.".
Blinding (performance bias and detection bias) All outcomes	Low risk	The researcher was blind, whilst there was not a feasible intervention to ensure the blinding of participants. Quote: "The researcher was blind to intervention allocation by the use of allocation lists and sealed envelopes containing the different interventions.". "Data entry and statistical analysis were completed by a researcher blind to intervention allocation." "it was not possible to claim that the participants were blind to the intervention level to which they were assigned as this would depend on each individual's familiarity with the current guidance."
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	17 CMHT were identified and 13 (76%) participated in the study; reasons for non participation were due to management concerns of time demands and no response from team members to recruitment. Of the 82 participants enrolled at the 13 CMHT, 68 (83%) returned both baseline and outcome data.
Selective reporting (reporting bias)	Low risk	All outcome measures reported in the Methods were described in the analysis and reported in the Results.
Other bias	Low risk	At baseline, both groups were equivalent in terms of demographic characteristics, intention scores to follow the guidance, and actual behaviours present consistent with the guidance.

Osborn 2010

Methods	Allocation: cluster-randomised trial.
	Blindness: not blind.
	Duration: six months.
Participants	Diagnosis: schizophrenia and other severe mental disorders (schizoaffective disorder, bipolar disorder, persistent delusional disorder, non-organic chronic psychosis).
	N of clusters: six community mental health teams (CMHTs).
	N of participants: 121.
	Sex: 66% M (experimental group); 55% M (control group).
	Mean age, years: 42.0 (experimental group); 43.1 (control group).
	Setting: Camden and Islington Mental Health and Social Care Trust, London.

Cochrane
Library

Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

Osborn 2010 *(Continued)*

Interventions	1. Nurse-led intervention plus education pack. The intervention lasted six months and targeted improving the levels of recording of cardiovascular risk factors required to estimate 10-year cardiovascular risk. The nurse monitored whether cardiovascular screening had occurred and sent prompts to primary and secondary care staff if screening had not occurred. The nurse offered screening herself to cover participants who had not received the complete battery of cardiovascular risk factor screenings. Target of the intervention: nurses.
	2. Education pack only: received copies of existing guidelines but no input by the nurse.
Outcomes	Practitioner impact: screened for cardiovascular risk factors (blood pressure, cholesterol, glucose, BMI, smoking status, Framingham score), as reported by participants and from GP notes.
Notes	Mean number of participants per cluster (m): 20.2.
	Intraclass coefficient (ICC): not reported.
	Design effect: 2.92.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "A statistician uninvolved in the trial randomly generated treatment allocation numbers".
Allocation concealment (selection bias)	Low risk	Quote: "Participating CMHTs were randomly allocated to each arm of the trial using a sealed envelope method. Once two CMHTs in each geographical area had agreed to participate, the envelopes were opened to determine the allocation for each CMHT".
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Quote: "It was not possible for the researcher, the CMHT workers or the patients to remain blind to allocation. The researcher could not remain blind to which arm the CMHTs had been allocated to when she assessed satisfaction with the intervention and who had performed screening (GP, CMHT worker or nurse). Therefore, during the evaluation phase, the possibility of observer bias was decreased by obtaining information on the main outcome directly from primary care clinical notes."
Incomplete outcome data (attrition bias) All outcomes	Low risk	115 of 121 participants were included in follow-up assessments
Selective reporting (reporting bias)	Low risk	All outcome measures reported in the Methods were included in the analysis and reported in the Results.
Other bias	Low risk	At baseline, the two groups were well balanced in terms of sociodemographic and clinical characteristics.
		Funding: The study was funded by a Trial Platform grant from the UK Medical Research Council. Reference: G0301032.

Thompson 2008

Methods	Allocation: cluster-randomised trial.
	Blindness: not blind.

Thompson 2008 *(Continued)*

	Duration: five months.
Participants	Diagnosis: severe mental disorders.
	N of clusters: 19 adult psychiatric units.
	N of participants: 480.
	Sex: 63.6% M (experimental group); 50.0% M (control group).
	Mean age, years: 42.2 (experimental group); 42.4 (control group).
	Setting: South-West England.
Interventions	1. Multifaceted implementation strategy. The first phase of the implementation strategy involved a 30-minute structured personal visit to consultant psychiatrists by a specially trained clinical psychiatric pharmacist. The structure of the visit was based on the social marketing principles of 'academic-detailing'. For the second part of the intervention, a workbook for both doctors and nurses was developed. This contained educational materials and specific cognitive techniques to challenge polypharmacy. The cognitive behavioural techniques were based on the principles of reducing risk-taking behaviours. The workbook was distributed to all ward doctors and nurses. A range of strategies were offered as alternatives to polypharmacy. A 'booster' pamphlet was sent eight weeks after distribution of the workbook. For the third part of the intervention, a medication chart reminder system was developed. Ward pharmacists applied removable reminder stickers to medication charts when participants were prescribed more than one antipsychotic. Target of the intervention: doctors and nurses.
	2. Passive guideline dissemination: disseminated to all doctors and nurses.
Outcomes	Practitioner impact: antipsychotic polypharmacy (from participants' medication charts).
Notes	Mean number of participants per cluster (m): 25.3.
	Intraclass coefficient (ICC): 0.027.
	Design effect: 1.65.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers were generated by one of the study authors using a calculator's random number generator function.
Allocation concealment (selection bias)	Low risk	Allocation was performed by one of the study authors, who was blind to the identity of the units.
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Quote: "Participants were potentially aware of which group (control or intervention) they were allocated to due to the nature of the intervention".
Incomplete outcome data (attrition bias) All outcomes	Low risk	All randomly assigned units (clusters) were included in the analysis. The study consisted of two cross-sectional surveys, one carried out at baseline and the second at follow-up. The primary outcome was polypharmacy at follow-up (cross-sectional nature of the primary outcome measure).
Selective reporting (reporting bias)	Low risk	All outcome measures reported in the Methods were included in the analysis and reported in the Results.
Other bias	Unclear risk	Quote: "Patients were younger, more likely to be male and detained under the Mental Health Act in the intervention arm".

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Azrin 1998	Allocation: randomised.
	Participants: included people with bipolar disorder and unipolar; depression, not schizophrenia.
Brunette 2015	Allocation: randomised.
	Participants: young people with first episode psychosis.
	Intervention: a program for first episode psychosis, versus standard care, not guideline implementation.
Byng 2004	Allocation: randomised.
	Participants: included people with bipolar disorder, chronic. depressive or neurotic conditions, not schizophrenia.
De Beurs 2013	Allocation: randomised.
	Participants: included patients of various diagnostic categories (personality disorder, depressive disorder, anxiety disorder, and psychotic disorder). Randomized patients had mainly diagnosis of depression and personality disorders.
Koivunen 2010	Allocation: randomised.
	Participants: people with schizophrenia.
	Intervention: Web-based participant support system versus standard care, not guideline implementation.
Robinson 2015	Allocation: not randomised.
Rush 2003	Allocation: not randomised; control clinics were matched with experimental clinics.
Tjia 2014	Allocation: randomised.
	Participants: nursing homes residents with dementia-related behaviours.

Characteristics of studies awaiting assessment *[ordered by study ID]*

JPRN-UMIN000004931

Methods	Allocation: randomised
Participants	Diagnosis: schizophrenia
Interventions	Algorithm-guided pharmacological treatment (unclear implementation strategy) versus treatment as usual
Outcomes	Primary Outcome Measure: Positive and Negative Syndrome Scale. Secondary Outcomes Measures: Clinical Global Impression Scale-Schizophrenia Version (CGI-SCH), Global Assessment of Functioning (GAF), Drug-Induced Extrapyramidal Symptoms Scale (DIEPSS), Targeted Inventory on Problem in Schizophrenia (TIP-Sz) and Assessment for Comprehensive Treatment of Schizophrenia

JPRN-UMIN000004931 *(Continued)*

	(FACT-Sz), Short-form 36 v2 Health Survey (SF-36v2), Subjective Wellbeing under Neuroleptic Treatment Scale (SWN-J)
Notes	The study is described as completed (upload.umin.ac.jp/cgi-open-bin/ctr/ctr.cgi?function=brows&action=brows&type=summary&recptno=R000005869&language=E), but no results were posted. We contacted the authors for further information.

NCT00156637

Methods	Allocation: randomised
Participants	Diagnosis: Schizophrenia, schizoaffective disorder
Interventions	Team-Based Quality Improvement Intervention versus Opinion Leader Intervention
Outcomes	Primary Outcome Measures: Antipsychotic dose above guideline-recommended range, Side effect monitoring 30 days before/after new antipsychotic, Clozapine prescribing for treatment-refractory schizophrenia
	Secondary Outcome Measures: Patient functional status, Improvement in schizophrenia symptoms, Service Use
Notes	The study is described as completed (clinicaltrials.gov/ct2/show/NCT00156637), but no results were posted. The two publications indicated for this study refer to other conditions (PTSD) or are not randomised. We contacted the authors for further information on these aspects.

Olfson 1998

Methods	Allocation: cluster-randomised, outpatient clinics matched by sociodemographic composition of patients.
Participants	Diagnosis: schizophrenia.
	N = unclear.
Interventions	1. Medication algorithms: intensive guideline implementation intervention aimed at improving the short-term outcome of public sector participants with schizophrenia.
	2. Participant and family education.
	3. Clinical support.
Outcomes	Psychiatric symptoms, adverse effects, adherence to prescribed medications, social and occupational functioning, quality of life, satisfaction with care, mental healthcare service utilisation; unclear how these were measured.
Notes	

Owen 2013

Methods	Allocation: cluster-randomised
Participants	Veteran Affairs Medical Centers (VA) employees who are involved in the monitoring and management of patients prescribed with a new antipsychotic medication.

Owen 2013 *(Continued)*

	N = unclear
Interventions	1. evidence-based quality improvement approach (EBQI)
	2. evidence-based quality improvement approach plus external facilitation (EBQI/F)
Outcomes	Performance on recommended management actions,
Notes	Published material about this study refers to the study protocol.
	The study recruited patients taking antipsychotics medication, but the diagnosis of the participants is not clear. Moreover, it does not specified if the study sites (Veteran Affairs Medical Centers) include mental healthcare settings. We contacted the authors for further information on these aspects.

Characteristics of ongoing studies *[ordered by study ID]*

NCT00182494

Trial name or title	NCT00182494
Methods	Allocation: randomised.
	Blinding: double.
Participants	Diagnosis: schizophrenia.
Interventions	1. Behaviour: modified diabetes prevention protocol and metformin.
Outcomes	Incidence of new-onset diabetes.
Starting date	Study start: February 2005; expected completion: July 2009.
	Last follow-up: January 2009; data entry closure: April 2009.
Contact information	Lakshmi P Voruganti, MD, Principal Investigator, Hamilton Health Sciences, Ontario, Canada.
Notes	In January 2016 the recruitment status of this study was unknown because the information has not been verified recently; no publications provided (clinicaltrials.gov/ct2/show/NCT001824941). We contacted the authors for clarifications.

D A T A   A N D   A N A L Y S E S

Comparison 1.   Active education + Support for implementation guidelines versus Routine care or Passive dissemination

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Process outcome, Practitioner Impact 1. Polypharmacy at follow-up	2		Risk Ratio (M-H, Fixed, 95% CI)	Subtotals only

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Uncorrected	2	1082	Risk Ratio (M-H, Fixed, 95% CI)	1.10 [0.99, 1.23]
1.2 Corrected for design effect	2	310	Risk Ratio (M-H, Fixed, 95% CI)	0.97 [0.75, 1.25]
2 Process outcome, Practitioner Impact: 2. Not screened for cardiovascular risk	1		Risk Ratio (M-H, Random, 95% CI)	Subtotals only
2.1 Blood pressure—uncorrected	1	96	Risk Ratio (M-H, Random, 95% CI)	0.07 [0.02, 0.28]
2.2 Blood pressure—corrected for design effect	1	33	Risk Ratio (M-H, Random, 95% CI)	0.10 [0.01, 0.74]
2.3 Cholesterol—uncorrected	1	103	Risk Ratio (M-H, Random, 95% CI)	0.46 [0.30, 0.70]
2.4 Cholesterol—corrected for design effect	1	35	Risk Ratio (M-H, Random, 95% CI)	0.49 [0.24, 0.99]
2.5 Glucose—uncorrected	1	103	Risk Ratio (M-H, Random, 95% CI)	0.53 [0.34, 0.82]
2.6 Glucose—corrected for design effect	1	35	Risk Ratio (M-H, Random, 95% CI)	0.58 [0.28, 1.21]
2.7 BMI—uncorrected	1	99	Risk Ratio (M-H, Random, 95% CI)	0.22 [0.08, 0.60]
2.8 BMI—corrected for design effect	1	34	Risk Ratio (M-H, Random, 95% CI)	0.18 [0.02, 1.37]
2.9 Smoking status—uncorrected	1	96	Risk Ratio (M-H, Random, 95% CI)	0.28 [0.12, 0.64]
2.10 Smoking status—corrected for design effect	1	32	Risk Ratio (M-H, Random, 95% CI)	0.25 [0.06, 1.03]
2.11 Framingham score—uncorrected	1	110	Risk Ratio (M-H, Random, 95% CI)	0.69 [0.55, 0.87]
2.12 Framingham score—corrected for design effect	1	38	Risk Ratio (M-H, Random, 95% CI)	0.71 [0.48, 1.03]
3 Patient outcome: global state: average score (PANSS total, high = poor)	1		Mean Difference (IV, Random, 95% CI)	Subtotals only
3.1 Uncorrected	1	105	Mean Difference (IV, Random, 95% CI)	-1.30 [-8.21, 5.61]
3.2 Corrected for design effect	1	59	Mean Difference (IV, Random, 95% CI)	-1.30 [-10.52, 7.92]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
4 Patient outcome: satisfaction with care: average score (ZUF8, high = better satisfaction)	1		Mean Difference (IV, Random, 95% CI)	Subtotals only
4.1 Uncorrected	1	83	Mean Difference (IV, Random, 95% CI)	0.10 [-1.43, 1.63]
4.2 Corrected for design effect	1	46	Mean Difference (IV, Random, 95% CI)	0.10 [-1.96, 2.16]
5 Patient outcome: treatment adherence: not adherent at follow-up	1		Risk Ratio (M-H, Random, 95% CI)	Subtotals only
5.1 Uncorrected	1	349	Risk Ratio (M-H, Random, 95% CI)	0.87 [0.66, 1.15]
5.2 Corrected for design effect	1	52	Risk Ratio (M-H, Random, 95% CI)	0.9 [0.44, 1.85]
6 Patient outcome: drug attitude: average score (DAI, high = positive attitude)	1		Mean Difference (IV, Random, 95% CI)	Subtotals only
6.1 Uncorrected	1	57	Mean Difference (IV, Random, 95% CI)	-1.40 [-2.88, 0.08]
6.2 Corrected for design effect	1	32	Mean Difference (IV, Random, 95% CI)	-1.40 [-3.38, 0.58]

Analysis 1.1. Comparison 1 Active education + Support for implementation guidelines versus Routine care or Passive dissemination, Outcome 1 Process outcome, Practitioner Impact 1. Polypharmacy at follow-up.

Study or subgroup	Experimental n/N	Control n/N	Risk Ratio M-H, Fixed, 95% CI	Weight	Risk Ratio M-H, Fixed, 95% CI
1.1.1 Uncorrected					
Baandrup 2010	155/216	234/386		62.75%	1.18[1.05,1.33]
Thompson 2008	105/260	92/220		37.25%	0.97[0.78,1.2]
Subtotal (95% CI)	**476**	**606**		**100%**	**1.1[0.99,1.23]**
Total events: 260 (Experimental), 326 (Control)					
Heterogeneity: Tau²=0; Chi²=2.9, df=1(P=0.09); I²=65.56%					
Test for overall effect: Z=1.8(P=0.07)					
1.1.2 Corrected for design effect					
Baandrup 2010	5/7	8/12		8.84%	1.07[0.58,1.98]
Thompson 2008	64/158	56/133		91.16%	0.96[0.73,1.27]
Subtotal (95% CI)	**165**	**145**		**100%**	**0.97[0.75,1.25]**
Total events: 69 (Experimental), 64 (Control)					
Heterogeneity: Tau²=0; Chi²=0.1, df=1(P=0.75); I²=0%					
Test for overall effect: Z=0.22(P=0.83)					
Test for subgroup differences: Chi²=0.8, df=1 (P=0.37), I²=0%					

Favours experimental   0.5   0.7   1   1.5   2   Favours control

Analysis 1.2. Comparison 1 Active education + Support for implementation guidelines versus Routine care or Passive dissemination, Outcome 2 Process outcome, Practitioner Impact: 2. Not screened for cardiovascular risk.

Study or subgroup	Experimental n/N	Control n/N	Risk Ratio M-H, Random, 95% CI	Weight	Risk Ratio M-H, Random, 95% CI
1.2.1 Blood pressure—uncorrected					
Osborn 2010	2/52	24/44		100%	0.07[0.02,0.28]
Subtotal (95% CI)	52	44		100%	0.07[0.02,0.28]
Total events: 2 (Experimental), 24 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=3.75(P=0)					
1.2.2 Blood pressure—corrected for design effect					
Osborn 2010	1/18	8/15		100%	0.1[0.01,0.74]
Subtotal (95% CI)	18	15		100%	0.1[0.01,0.74]
Total events: 1 (Experimental), 8 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=2.26(P=0.02)					
1.2.3 Cholesterol—uncorrected					
Osborn 2010	17/51	38/52		100%	0.46[0.3,0.7]
Subtotal (95% CI)	51	52		100%	0.46[0.3,0.7]
Total events: 17 (Experimental), 38 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=3.65(P=0)					
1.2.4 Cholesterol—corrected for design effect					
Osborn 2010	6/17	13/18		100%	0.49[0.24,0.99]
Subtotal (95% CI)	17	18		100%	0.49[0.24,0.99]
Total events: 6 (Experimental), 13 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=1.99(P=0.05)					
1.2.5 Glucose—uncorrected					
Osborn 2010	17/51	33/52		100%	0.53[0.34,0.82]
Subtotal (95% CI)	51	52		100%	0.53[0.34,0.82]
Total events: 17 (Experimental), 33 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=2.87(P=0)					
1.2.6 Glucose—corrected for design effect					
Osborn 2010	6/17	11/18		100%	0.58[0.28,1.21]
Subtotal (95% CI)	17	18		100%	0.58[0.28,1.21]
Total events: 6 (Experimental), 11 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=1.45(P=0.15)					
1.2.7 BMI—uncorrected					
Osborn 2010	4/53	16/46		100%	0.22[0.08,0.6]
Subtotal (95% CI)	53	46		100%	0.22[0.08,0.6]
Total events: 4 (Experimental), 16 (Control)					
Heterogeneity: Not applicable					

Favours experimental 0.005 0.1 1 10 200 Favours control

Study or subgroup	Experimental n/N	Control n/N	Risk Ratio M-H, Random, 95% CI	Weight	Risk Ratio M-H, Random, 95% CI
Test for overall effect: Z=2.93(P=0)					
1.2.8 BMI—corrected for design effect					
Osborn 2010	1/18	5/16		100%	0.18[0.02,1.37]
Subtotal (95% CI)	**18**	**16**		**100%**	**0.18[0.02,1.37]**
Total events: 1 (Experimental), 5 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=1.66(P=0.1)					
1.2.9 Smoking status—uncorrected					
Osborn 2010	6/51	19/45		100%	0.28[0.12,0.64]
Subtotal (95% CI)	**51**	**45**		**100%**	**0.28[0.12,0.64]**
Total events: 6 (Experimental), 19 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=3.03(P=0)					
1.2.10 Smoking status—corrected for design effect					
Osborn 2010	2/17	7/15		100%	0.25[0.06,1.03]
Subtotal (95% CI)	**17**	**15**		**100%**	**0.25[0.06,1.03]**
Total events: 2 (Experimental), 7 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=1.92(P=0.06)					
1.2.11 Framingham score—uncorrected					
Osborn 2010	34/55	49/55		100%	0.69[0.55,0.87]
Subtotal (95% CI)	**55**	**55**		**100%**	**0.69[0.55,0.87]**
Total events: 34 (Experimental), 49 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=3.15(P=0)					
1.2.12 Framingham score—corrected for design effect					
Osborn 2010	12/19	17/19		100%	0.71[0.48,1.03]
Subtotal (95% CI)	**19**	**19**		**100%**	**0.71[0.48,1.03]**
Total events: 12 (Experimental), 17 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=1.81(P=0.07)					

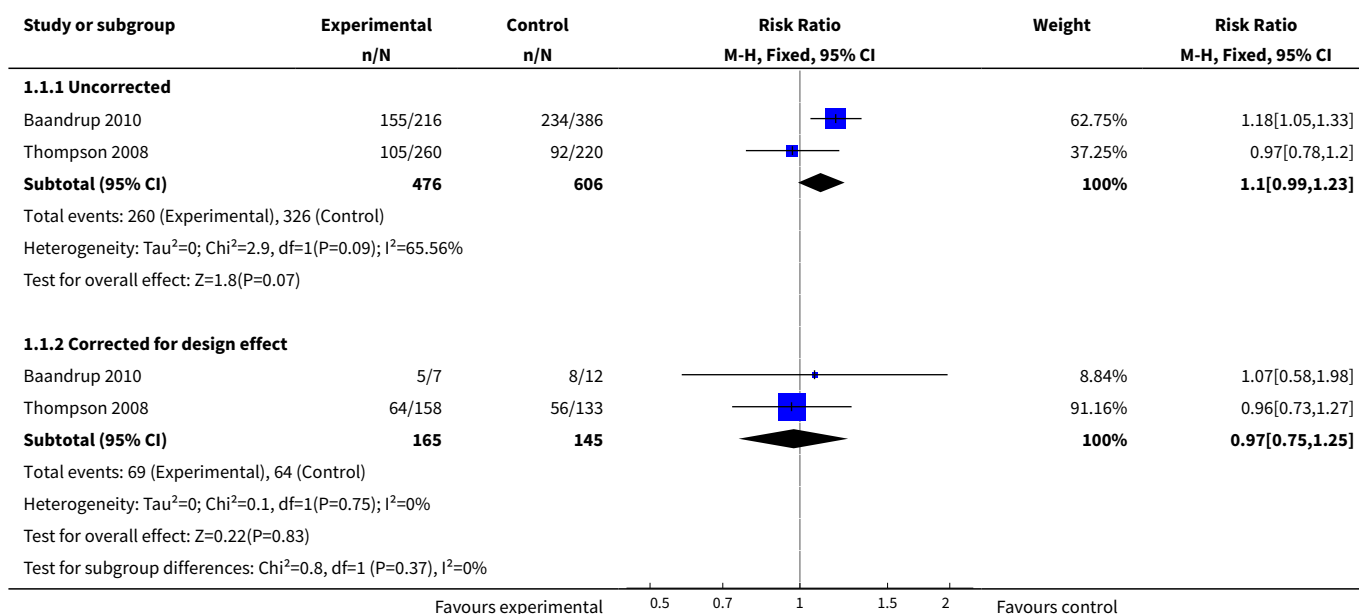Favours experimental   0.005   0.1   1   10   200   Favours control

Analysis 1.3. Comparison 1 Active education + Support for implementation guidelines versus Routine care or Passive dissemination, Outcome 3 Patient outcome: global state: average score (PANSS total, high = poor).

Study or subgroup	Experimental N	Mean(SD)	Control N	Mean(SD)	Mean Difference Random, 95% CI	Weight	Mean Difference Random, 95% CI
1.3.1 Uncorrected							
Hamann 2006	48	58 (17.3)	57	59.3 (18.8)		100%	-1.3[-8.21,5.61]
Subtotal ***	**48**		**57**			**100%**	**-1.3[-8.21,5.61]**
Heterogeneity: Not applicable							
Test for overall effect: Z=0.37(P=0.71)							

Favours experimental   -20   -10   0   10   20   Favours control

Study or subgroup	Experimental		Control		Mean Difference	Weight	Mean Difference
	N	Mean(SD)	N	Mean(SD)	Random, 95% CI		Random, 95% CI
1.3.2 Corrected for design effect							
Hamann 2006	27	58 (17.3)	32	59.3 (18.8)		100%	-1.3[-10.52,7.92]
Subtotal ***	**27**		**32**			**100%**	**-1.3[-10.52,7.92]**
Heterogeneity: Not applicable							
Test for overall effect: Z=0.28(P=0.78)							
Test for subgroup differences: Not applicable							

Favours experimental -20 -10 0 10 20 Favours control

Analysis 1.4. Comparison 1 Active education + Support for implementation guidelines versus Routine care or Passive dissemination, Outcome 4 Patient outcome: satisfaction with care: average score (ZUF8, high = better satisfaction).

Study or subgroup	Experimental		Control		Mean Difference	Weight	Mean Difference
	N	Mean(SD)	N	Mean(SD)	Random, 95% CI		Random, 95% CI
1.4.1 Uncorrected							
Hamann 2006	34	-16.3 (3.7)	49	-16.4 (3.2)		100%	0.1[-1.43,1.63]
Subtotal ***	**34**		**49**			**100%**	**0.1[-1.43,1.63]**
Heterogeneity: Tau²=0; Chi²=0, df=0(P<0.0001); I²=100%							
Test for overall effect: Z=0.13(P=0.9)							
1.4.2 Corrected for design effect							
Hamann 2006	19	-16.3 (3.7)	27	-16.4 (3.2)		100%	0.1[-1.96,2.16]
Subtotal ***	**19**		**27**			**100%**	**0.1[-1.96,2.16]**
Heterogeneity: Not applicable							
Test for overall effect: Z=0.1(P=0.92)							
Test for subgroup differences: Chi²=0, df=1 (P=1), I²=0%							

Favours experimental -5 -2.5 0 2.5 5 Favours control

Analysis 1.5. Comparison 1 Active education + Support for implementation guidelines versus Routine care or Passive dissemination, Outcome 5 Patient outcome: treatment adherence: not adherent at follow-up.

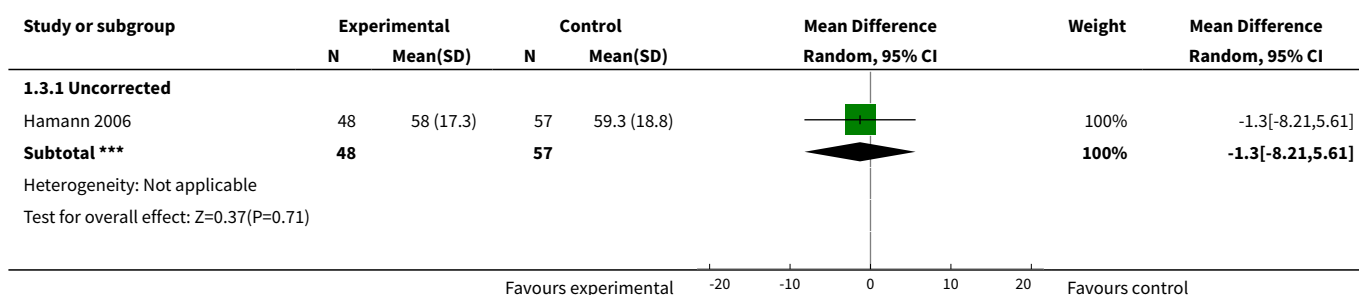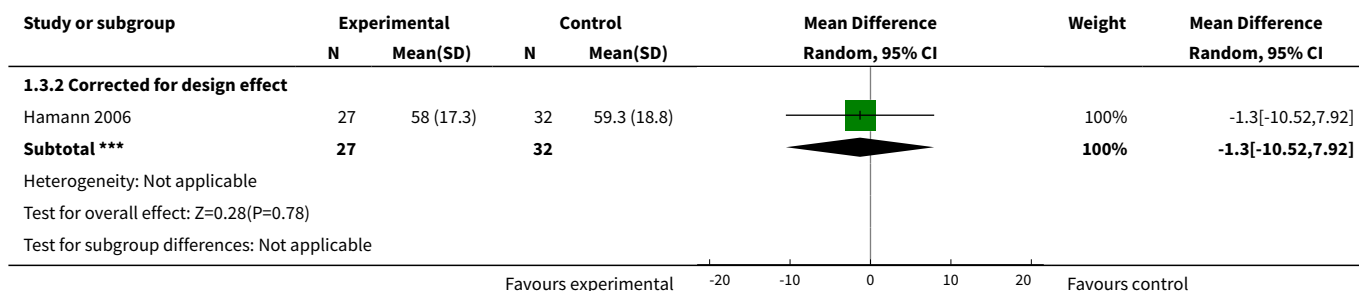| Study or subgroup	Experimental	Control	Risk Ratio	Weight	Risk Ratio
	n/N	n/N	M-H, Random, 95% CI		M-H, Random, 95% CI
1.5.1 Uncorrected					
Hudson 2008	60/173	70/176		100%	0.87[0.66,1.15]
Subtotal (95% CI)	**173**	**176**		**100%**	**0.87[0.66,1.15]**
Total events: 60 (Experimental), 70 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=0.98(P=0.33)					
1.5.2 Corrected for design effect					
Hudson 2008	9/26	10/26		100%	0.9[0.44,1.85]
Subtotal (95% CI)	**26**	**26**		**100%**	**0.9[0.44,1.85]**
Total events: 9 (Experimental), 10 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=0.29(P=0.77)					

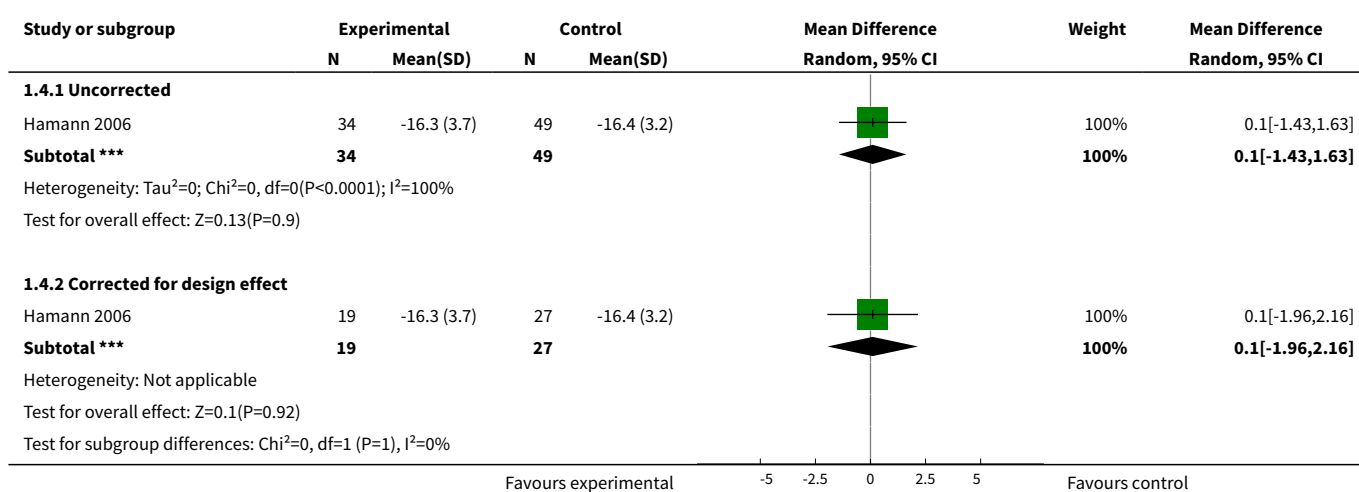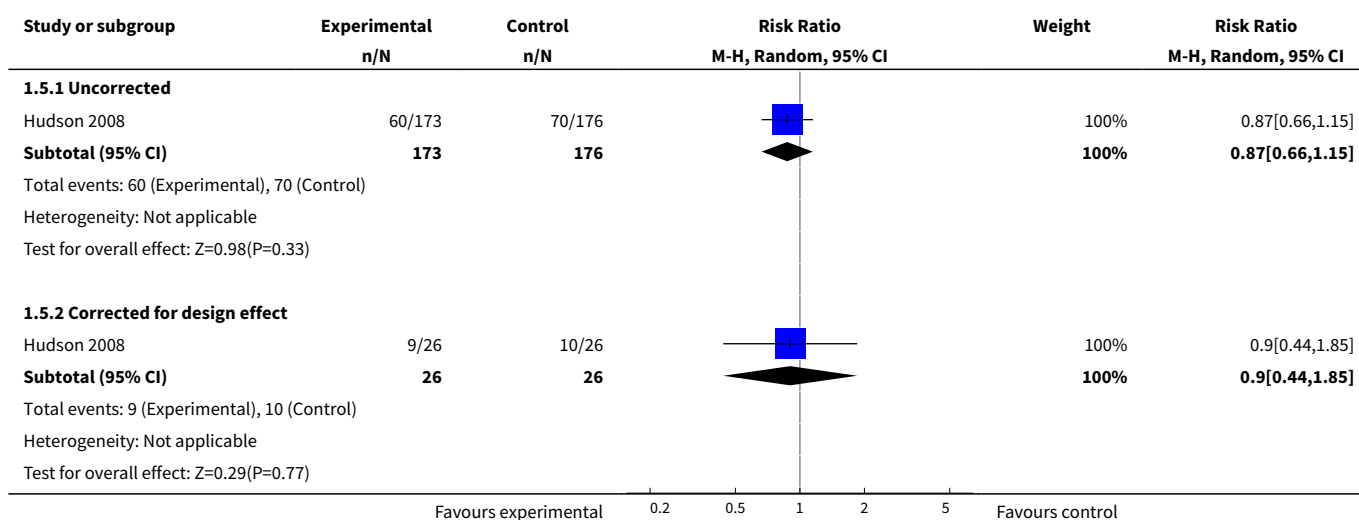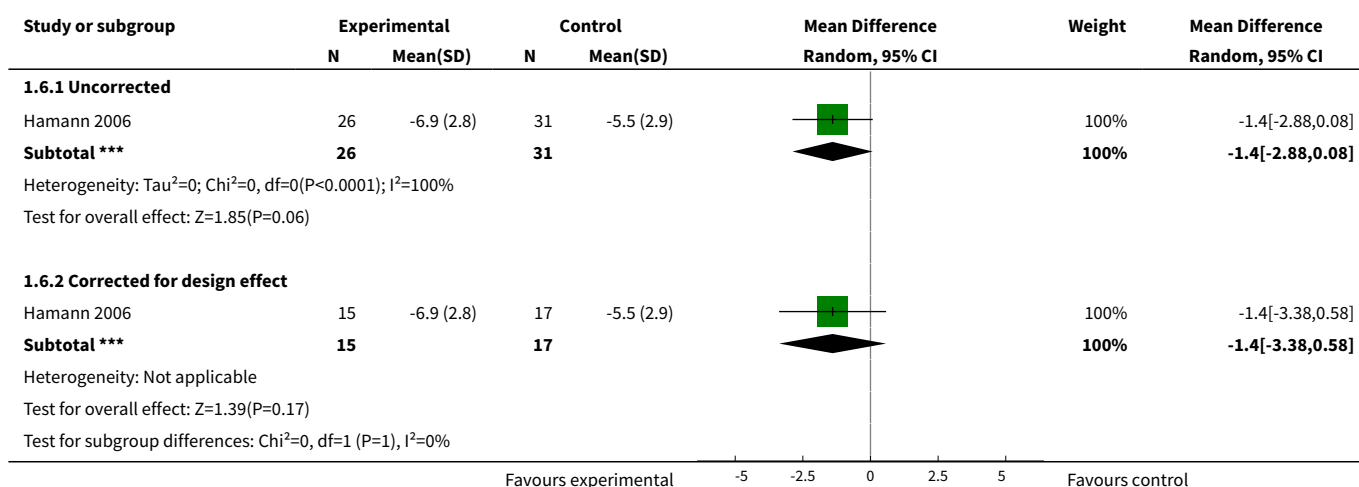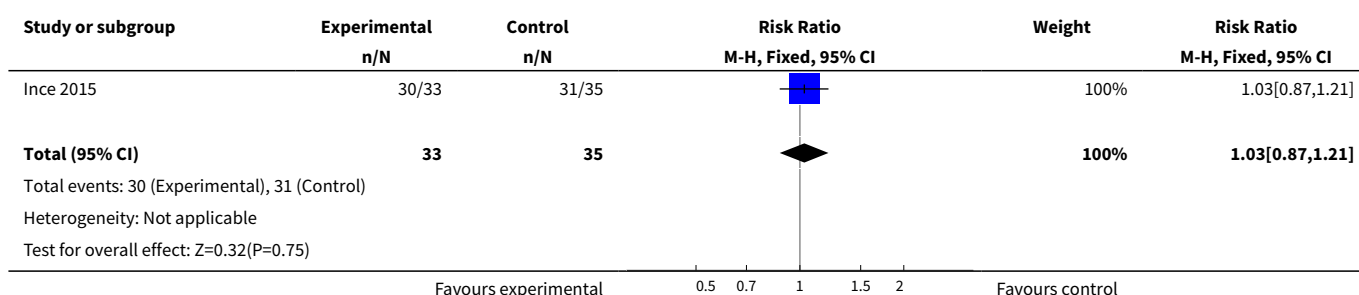Favours experimental 0.2 0.5 1 2 5 Favours control

Analysis 1.6. Comparison 1 Active education + Support for implementation guidelines versus Routine care or Passive dissemination, Outcome 6 Patient outcome: drug attitude: average score (DAI, high = positive attitude).
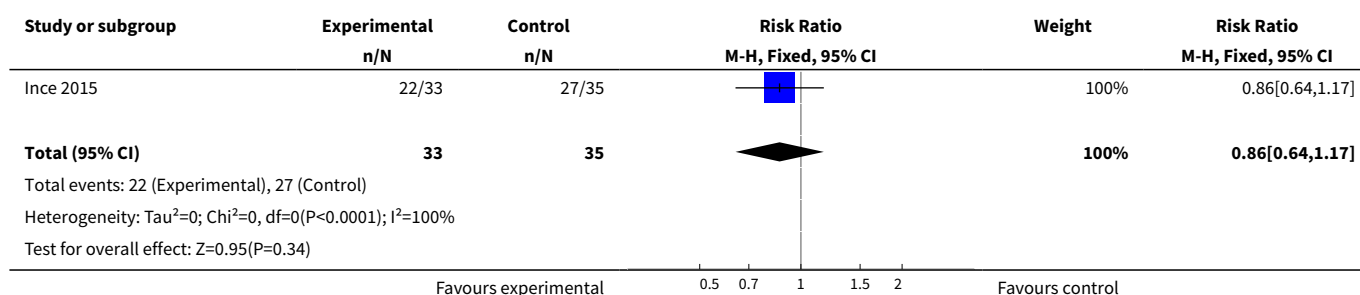
Study or subgroup	Experimental		Control		Mean Difference	Weight	Mean Difference
	N	Mean(SD)	N	Mean(SD)	Random, 95% CI		Random, 95% CI
1.6.1 Uncorrected							
Hamann 2006	26	-6.9 (2.8)	31	-5.5 (2.9)		100%	-1.4[-2.88,0.08]
Subtotal ***	**26**		**31**			**100%**	**-1.4[-2.88,0.08]**
Heterogeneity: Tau²=0; Chi²=0, df=0(P<0.0001); I²=100%							
Test for overall effect: Z=1.85(P=0.06)							
1.6.2 Corrected for design effect							
Hamann 2006	15	-6.9 (2.8)	17	-5.5 (2.9)		100%	-1.4[-3.38,0.58]
Subtotal ***	**15**		**17**			**100%**	**-1.4[-3.38,0.58]**
Heterogeneity: Not applicable							
Test for overall effect: Z=1.39(P=0.17)							
Test for subgroup differences: Chi²=0, df=1 (P=1), I²=0%							

Favours experimental -5 -2.5 0 2.5 5 Favours control

Comparison 2. Use of re-written guideline text versus use of original guideline text

Outcome or subgroup title	No. of studies	No. of partici-pants	Statistical method	Effect size
1 Process outcome, Practitioner Impact 3. Staff not trained (as suggested by the guideline)	1	68	Risk Ratio (M-H, Fixed, 95% CI)	1.03 [0.87, 1.21]
2 Process outcome, Practitioner Impact 4. Staff not supervised (as suggested by the guideline)	1	68	Risk Ratio (M-H, Fixed, 95% CI)	0.86 [0.64, 1.17]
3 Process outcome, Practitioner Impact 5. Staff did not provide psychological interventions (as suggested by the guideline)	1	68	Risk Ratio (M-H, Fixed, 95% CI)	0.86 [0.62, 1.18]

Analysis 2.1. Comparison 2 Use of re-written guideline text versus use of original guideline text, Outcome 1 Process outcome, Practitioner Impact 3. Staff not trained (as suggested by the guideline).

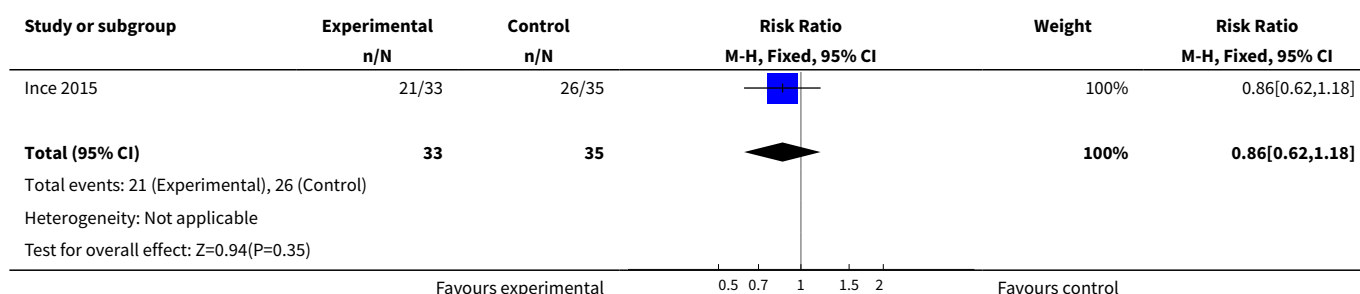| Study or subgroup	Experimental	Control	Risk Ratio	Weight	Risk Ratio
	n/N	n/N	M-H, Fixed, 95% CI		M-H, Fixed, 95% CI
Ince 2015	30/33	31/35		100%	1.03[0.87,1.21]
Total (95% CI)	**33**	**35**		**100%**	**1.03[0.87,1.21]**
Total events: 30 (Experimental), 31 (Control)					
Heterogeneity: Not applicable					
Test for overall effect: Z=0.32(P=0.75)					

Favours experimental 0.5 0.7 1 1.5 2 Favours control

Analysis 2.2. Comparison 2 Use of re-written guideline text versus use of original guideline text, Outcome 2 Process outcome, Practitioner Impact 4. Staff not supervised (as suggested by the guideline).

Study or subgroup	Experimental n/N	Control n/N	Risk Ratio M-H, Fixed, 95% CI	Weight	Risk Ratio M-H, Fixed, 95% CI
Ince 2015	22/33	27/35		100%	0.86[0.64,1.17]
Total (95% CI)	**33**	**35**		**100%**	**0.86[0.64,1.17]**

Total events: 22 (Experimental), 27 (Control)

Heterogeneity: Tau²=0; Chi²=0, df=0(P<0.0001); I²=100%

Test for overall effect: Z=0.95(P=0.34)

Favours experimental 0.5 0.7 1 1.5 2 Favours control

Analysis 2.3. Comparison 2 Use of re-written guideline text versus use of original guideline text, Outcome 3 Process outcome, Practitioner Impact 5. Staff did not provide psychological interventions (as suggested by the guideline).

Study or subgroup	Experimental n/N	Control n/N	Risk Ratio M-H, Fixed, 95% CI	Weight	Risk Ratio M-H, Fixed, 95% CI
Ince 2015	21/33	26/35		100%	0.86[0.62,1.18]
Total (95% CI)	**33**	**35**		**100%**	**0.86[0.62,1.18]**

Total events: 21 (Experimental), 26 (Control)

Heterogeneity: Not applicable

Test for overall effect: Z=0.94(P=0.35)

Favours experimental 0.5 0.7 1 1.5 2 Favours control

ADDITIONAL TABLES

Table 1. Suggested schizophrenia reviews into which excluded randomised studies could be used

Excluded study tag	Comparison	Existing Cochrane review
Koivunen 2010	Web-based participant support systems for people with schizophrenia.	Välimäki 2012

Table 2. Suggested design for future trial

Methods	Allocation: random allocation—As the intervention naturally occurs in groups of individuals, the unit of random allocation may be the group, or cluster, rather than the individual.
	Concealment of cluster allocation status: Participant recruitment may be carried out by professionals who are masked to the cluster allocation status.
	Blinding: Outcome assessors should not be involved in the conduct of the study and may be masked to the allocated interventions.
Participants	Diagnosis: people with schizophrenia or related psychosis (clinicians' implicit criteria)
	Sample size: depends on number of clusters, average cluster size, and degree of correlation within clusters (intracluster correlation coefficient (ICC)).

Table 2. Suggested design for future trial *(Continued)*

| Interventions | 1. Guideline implementation strategy. |
| | 2. Passive dissemination. |

| Outcomes | Primary outcomes: practitioner impact (this measure would differ according to the characteristics and purposes of the guideline under trial); participant impact (global state, this measure would differ according to the characteristics and purposes of the guideline under trial). |
| | Secondary outcomes: treatment adherence, satisfaction with care, drug attitude, and quality of life. |

| Note | Cluster trials should report design effect and ICC. |

WHAT'S NEW

Date	Event	Description
18 February 2016	New search has been performed	Text and analysis of the review updated with the newly included study
21 December 2015	New citation required but conclusions have not changed	Results from 2015 search added to review; 1 new study included
25 August 2015	Amended	Search was updated and 103 new references were added to 'Classification pending references' section of the review.

CONTRIBUTIONS OF AUTHORS

2016 update

Irene Bighelli - study identification, critical appraisal, data entry, interpretation, and writing.

Giovanni Ostuzzi - study identification, critical appraisal, and data entry.

Corrado Barbui - interpretation and writing.

Previous version

Corrado Barbui - study identification, critical appraisal, data entry, interpretation, and writing.

Francesca Girlanda - study identification, critical appraisal, and data entry.

Esra Ay - study identification and critical appraisal.

Andrea Cipriani - interpretation and writing.

Thomas Becker - interpretation and writing.

Markus Koesters - interpretation and writing.

DECLARATIONS OF INTEREST

Irene Bighelli - none known.

Giovanni Ostuzzi - none known.

Corrado Barbui - none known.

Francesca Girlanda - none known.

Esra Ay - none known.

Andrea Cipriani - was expert witness for Accord Healthcare for a patent issue about quetiapine extended release.

Thomas Becker - author's Department has received funding from the German Federal Ministry for Education and Research (BMBF) for a systematic review of the effects of guideline implementation in the treatment of people with mental disorders. Thomas Becker was lead of the author team for a (S3) guideline of the Deutsche Gesellschaft für Psychiatrie, Psychotherapie, Psychosomatik und Nervenheilkunde (DGPPN) on psychosocial interventions in people with severe mental disorders.

Markus Koesters - none known.

SOURCES OF SUPPORT

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

At the protocol stage, the 'Summary of findings' outcomes were not listed; in the review, we specified that the 'Summary of findings' outcomes correspond to the primary and secondary outcomes of the review and that we would have preferred these data to be binary for the purposes of the table.

INDEX TERMS

Medical Subject Headings (MeSH)

*Guideline Adherence; *Outcome and Process Assessment, Health Care; *Practice Guidelines as Topic; *Specialization; Antipsychotic Agents [*therapeutic use]; Mental Health; Randomized Controlled Trials as Topic; Schizophrenia [*drug therapy]

MeSH check words

Humans